

## CHEMICAL INHIBITORS OF MISMATCH REPAIR

### TECHNICAL FIELD OF THE INVENTION

The invention is related to the area of mutagenesis. In particular it is related to the field of blocking specific DNA repair processes.

### BACKGROUND OF THE INVENTION

Mismatch repair (MMR) is a conserved DNA repair process that is involved in post-replicative repair of mutated DNA sequences that occurs after genome replication. The process involves a group of gene products, including the *mutS* homologs *GTBP*, *hMSH2*, and *hMSH3* and the *mutL* homologs *hMLH1*, *hPMS1*, and *hPMS2* (Bronner, C.E. *et al.* (1994) *Nature* 368:258-261; Papadopoulos, N. *et al.* (1994) *Science* 263:1625-1629; Leach, F.S. *et al.* (1993) *Cell* 75:1215-1225; Nicolaides, N.C. *et al.* (1994) *Nature* 371:75-80) that work in concert to correct mispaired mono-, di-, and tri-nucleotides, point mutations, and to monitor for correct homologous recombination. Germline mutations in any of the genes involved in this process results in global point mutations, and instability of mono, di and tri-nucleotide repeats (a feature referred to as microsatellite instability (MI)), throughout the genome of the host cell. In man, genetic defects in MMR results in the predisposition to hereditary nonpolyposis colon cancer, a disease in which tumors retain a diploid genome but have widespread MI (Bronner, C.E. *et al.* (1994) *Nature* 368:258-261; Papadopoulos, N. *et al.* (1994) *Science* 263:1625-1629; Leach, F.S. *et al.* (1993) *Cell* 75:1215-1225; Nicolaides, N.C. *et al.* (1994) *Nature* 371:75-80; Harfe B.D., and S. Jinks-Robertson (2000) *An. Rev. Genet.* 34:359-399; Modrich, P. (1994) *Science* 266:1959-1960). Though the mutator defect that arises from MMR deficiency can affect any DNA sequence, microsatellite sequences are particularly sensitive to MMR abnormalities (Peinado, M.A. *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:10065-10069). Microsatellite instability is therefore a useful indicator of defective MMR. In addition to its occurrence in virtually all tumors arising in HNPCC patients, MI is found in a small fraction of sporadic tumors with distinctive molecular and

phenotypic properties that is due to defective MMR (Perucho, M. (1996) *Biol. Chem.* 377:675-684).

MMR deficiency leads to a wide spectrum of mutations (point mutations, insertions, deletions, recombination, etc.) that can occur throughout the genome of a host cell. This effect has been found to occur across a diverse array of organisms ranging from but not limited to unicellular microbes, such as bacteria and yeast, to more complex organisms such as *Drosophila* and mammals, including mice and humans (Harfe B.D., and S. Jinks-Robertson (2000) *An. Rev. Genet.* 34:359-399; Modrich, P. (1994) *Science* 266:1959-1960). The ability to block MMR in a normal host cell or organism can result in the generation of genetically altered offspring or sibling cells that have desirable output traits for applications such as but not limited to agriculture, pharmaceutical, chemical manufacturing and specialty goods. A chemical method that can block the MMR process is beneficial for generating genetically altered hosts with commercially valuable output traits. A chemical strategy for blocking MMR *in vivo* offers a great advantage over a recombinant approach for producing genetically altered host organisms. One advantage is that a chemical approach bypasses the need for introducing foreign DNA into a host, resulting in a rapid approach for inactivating MMR and generating genetically diverse offspring or sib cells. Moreover, a chemical process is highly regulated in that once a host organism with a desired output trait is generated, the chemical is removed from the host and its MMR process would be restored, thus fixing the genetic alteration in subsequent generations. The invention described herein is directed to the discovery of small molecules that are capable of blocking MMR, thus resulting in host organisms with MI, a hallmark of MMR deficiency (Peinado, M.A. *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:10065-10069; Perucho, M. (1996) *Biol. Chem.* 377:675-684; Wheeler, J.M. *et al.* (2000) *J. Med. Genet.* 37:588-592; Hoang, J.M. *et al.* (1997) *Cancer Res.* 57:300-303). Moreover, host organisms exhibiting MI are then selected for to identify subtypes with new output traits, such as but not limited to mutant nucleic acid molecules, polypeptides, biochemicals, physical appearance at the microscopic and/or macroscopic level, or phenotypic alterations in a whole organism. In addition, the ability to develop MMR defective host cells by a chemical agent provides a valuable method for creating genetically altered cell hosts for product development. The invention described herein is directed to the

creation of genetically altered cell hosts via the blockade of MMR using chemical agents *in vivo*.

The advantages of the present invention are further described in the examples and figures described within this document.

### SUMMARY OF THE INVENTION

The invention provides methods for rendering cells hypermutable by blocking MMR activity with chemical agents.

The invention also provides genetically altered cell lines which have mutations introduced through interruption of mismatch repair.

The invention further provides methods to produce an enhanced rate of genetic hypermutation in a cell.

The invention encompasses methods of mutating a gene of interest in a cell, methods of creating cells with new phenotypes, and methods of creating cells with new phenotypes and a stable genome.

The invention also provides methods of creating genetically altered whole organisms and methods of creating whole organisms with new phenotypes.

These and other objects of the invention are provided by one or more of the embodiments described below.

In one embodiment of the invention, a method for screening chemical compounds that block mismatch repair (MMR) is provided. An MMR-sensitive reporter gene containing an out-of-frame polynucleotide repeat in its coding region is introduced into an MMR proficient cell. The cell is grown in the presence of chemicals. Chemicals that alter the genetic structure of the polynucleotide repeat yield a biologically active reporter gene product. Chemicals that disrupt the polynucleotide repeat are identified as MMR blocking agents.

In another embodiment of the invention, an isolated MMR blocking chemical is provided. The chemical can block MMR of a host cell, yielding a cell that exhibits an enhanced rate of hypermutation.

In another embodiment of the invention, a method is provided for introducing a mutation into a gene of interest. A chemical that blocks mismatch repair is added to the

culture of a cell line. The cells become hypermutable as a result of the introduction of the chemical. The cell further comprises a gene of interest. The cell is cultured and tested to determine whether the gene of interest harbors a mutation.

In another embodiment of the invention, a method is provided for producing new phenotypes of a cell. A chemical that blocks mismatch repair is added to a cell culture. The cell becomes hypermutable as a result of the introduction of the chemical. The cell is cultured and tested for the expression of new phenotypes.

In another embodiment of the invention, a method is provided for restoring genetic stability in a cell in which mismatch repair is blocked via a chemical agent. The chemical is removed from the cell culture and the cell restores its genetic stability.

In another embodiment of the invention, a method is provided for restoring genetic stability in a cell with blocked mismatch repair and a newly selected phenotype. The chemical agent is removed from the cell culture and the cell restores its genetic stability and the new phenotype is stable.

In another embodiment of the invention, a chemical method for blocking MMR in plants is provided. The plant is grown in the presence of a chemical agent. The plant is grown and exhibits an enhanced rate of hypermutation.

In another embodiment of the invention, a method for screening chemical inhibitors of MMR in plants *in vivo* is provided. MMR-sensitive plant expression vectors are engineered. The reporter vectors are introduced into plant hosts. The plant is grown in the presence of a chemical agent. The plant is monitored for altered reporter gene function.

In another embodiment of the invention, a method is provided for introducing a mutation into a gene of interest in a plant. A chemical that blocks mismatch repair is added to a plant. The plant becomes hypermutable as a result of the introduction of the chemical. The plant further comprises a gene of interest. The plant is grown. The plant is tested to determine whether the gene of interest harbors a mutation.

In another embodiment of the invention, a method is provided for producing new phenotypes of a plant. A chemical that blocks mismatch repair is added to a plant. The plant becomes hypermutable as a result of the introduction of the chemical. The plant is grown and tested for the expression of new phenotypes.

In another embodiment of the invention, a method is provided for restoring genetic stability in a plant in which mismatch repair is blocked via a chemical agent. The chemical is removed from the plant culture and the plant restores its genetic stability.

In another embodiment of the invention, a method is provided for restoring genetic stability in a plant with blocked mismatch repair and a newly selected phenotype. The chemical agent is removed from the plant culture and the plant restores its genetic stability and the new phenotype is stable.

These and other embodiments of the invention provide the art with methods that can generate enhanced mutability in microbes, organisms of the protista class, insect cells, mammalian cells, plants, and animals as well as providing cells, plants and animals harboring potentially useful mutations.

#### BRIEF DESCRIPTION OF THE DRAWINGS

**Figure 1** shows diagrams of mismatch repair (MMR) sensitive reporter genes.

**Figure 2** shows a screening method for identifying MMR blocking chemicals.

**Figure 3** shows identification of a small chemical that blocks MMR and genetically alters the pCAR-OF vector *in vivo*.

**Figure 4** shows shifting of endogenous microsatellites in human cells induced by a chemical inhibitor of MMR.

**Figure 5** shows sequence analysis of microsatellites from cells treated with chemical inhibitors of MMR with altered repeats.

**Figure 6** shows generation of host organisms with new phenotypes using a chemical blocker of MMR.

**Figure 7** shows a schematic diagram of MMR-sensitive reporter gene for plants.

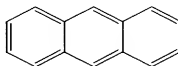
**Figure 8** shows derivatives of lead compounds and thereof that are inhibitors of MMR *in vivo*.

#### DETAILED DESCRIPTION OF THE INVENTION

Various definitions are provided herein. Most words and terms have the meaning that would be attributed to those words by one skilled in the art. Words or terms specifically

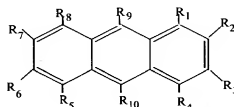
defined herein have the meaning provided in the context of the present invention as a whole and as are typically understood by those skilled in the art. Any conflict between an art-understood definition of a word or term and a definition of the word or term as specifically taught herein shall be resolved in favor of the latter. Headings used herein are for convenience and are not to be construed as limiting.

As used herein the term "anthracene" refers to the compound anthracene. However, when referred to in the general sense, such as "anthracenes," "an anthracene" or "the anthracene," such terms denote any compound that contains the fused triphenyl core structure of anthracene, i.e.,



regardless of extent of substitution.

In certain preferred embodiments of the invention, the anthracene has the formula:



wherein  $R_1$ - $R_{10}$  are independently hydrogen, hydroxyl, amino, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkynyl, substituted alkynyl, O-alkyl, S-alkyl, N-alkyl, O-alkenyl, S-alkenyl, N-alkenyl, O-alkynyl, S-alkynyl, N-alkynyl, aryl, substituted aryl, aryloxy, substituted aryloxy, heteroaryl, substituted heteroaryl, aralkyloxy, arylalkyl, alkylaryl, alkylaryloxy, arylsulfonyl, alkylsulfonyl, alkoxycarbonyl, aryloxy carbonyl, guanidino, carboxy, an alcohol, an amino acid, sulfonate, alkyl sulfonate, CN,  $\text{NO}_2$ , an aldehyde group, an ester, an ether, a crown ether, a ketone, an organosulfur compound, an organometallic group, a carboxylic acid, an organosilicon or a carbohydrate that optionally contains one or more alkylated hydroxyl groups;

wherein said heteroalkyl, heteroaryl, and substituted heteroaryl contain at least one heteroatom that is oxygen, sulfur, a metal atom, phosphorus, silicon or nitrogen; and

wherein said substituents of said substituted alkyl, substituted alkenyl, substituted alkynyl,

substituted aryl, and substituted heteroaryl are halogen, CN, NO<sub>2</sub>, lower alkyl, aryl, heteroaryl, aralkyl, aralkyloxy, guanidino, alkoxycarbonyl, alkoxy, hydroxy, carboxy and amino;

and wherein said amino groups optionally substituted with an acyl group, or 1 to 3 aryl or lower alkyl groups;

or wherein any two of R<sub>1</sub>-R<sub>10</sub> can together form a polyether;

or wherein any two of R<sub>1</sub>-R<sub>10</sub> can, together with the intervening carbon atoms of the anthracene core, form a crown ether.

As used herein, "alkyl" refers to a hydrocarbon containing from 1 to about 20 carbon atoms. Alkyl groups may straight, branched, cyclic, or combinations thereof. Alkyl groups thus include, by way of illustration only, methyl, ethyl, propyl, isopropyl, butyl, isobutyl, cyclopentyl, cyclopentylmethyl, cyclohexyl, cyclohexylmethyl, and the like. Also included within the definition of "alkyl" are fused and/or polycyclic aliphatic cyclic ring systems such as, for example, adamantane. As used herein the term "alkenyl" denotes an alkyl group having at least one carbon-carbon double bond. As used herein the term "alkynyl" denotes an alkyl group having at least one carbon-carbon triple bond.

In some preferred embodiments, the alkyl, alkenyl, alkynyl, aryl, aryloxy, and heteroaryl substituent groups described above may bear one or more further substituent groups; that is, they may be "substituted". In some preferred embodiments these substituent groups can include halogens (for example fluorine, chlorine, bromine and iodine), CN, NO<sub>2</sub>, lower alkyl groups, aryl groups, heteroaryl groups, aralkyl groups, aralkyloxy groups, guanidino, alkoxycarbonyl, alkoxy, hydroxy, carboxy and amino groups. In addition, the alkyl and aryl portions of aralkyloxy, arylalkyl, arylsulfonyl, alkylsulfonyl, alkoxycarbonyl, and aryloxy carbonyl groups also can bear such substituent groups. Thus, by way of example only, substituted alkyl groups include, for example, alkyl groups fluoro-, chloro-, bromo- and iodoalkyl groups, aminoalkyl groups, and hydroxyalkyl groups, such as hydroxymethyl, hydroxyethyl, hydroxypropyl, hydroxybutyl, and the like. In some preferred embodiments such hydroxyalkyl groups contain from 1 to about 20 carbons.

As used herein the term "aryl" means a group having 5 to about 20 carbon atoms and

which contains at least one aromatic ring, such as phenyl, biphenyl and naphthyl. Preferred aryl groups include unsubstituted or substituted phenyl and naphthyl groups. The term "aryloxy" denotes an aryl group that is bound through an oxygen atom, for example a phenoxy group.

In general, the prefix "hetero" denotes the presence of at least one hetero (i.e., non-carbon) atom, which is in some preferred embodiments independently one to three O, N, S, P, Si or metal atoms. Thus, the term "heteroaryl" denotes an aryl group in which one or more ring carbon atom is replaced by such a heteroatom. Preferred heteroaryl groups include pyridyl, pyrimidyl, pyrrolyl, furyl, thienyl, and imidazolyl groups.

The term "aralkyl" (or "arylalkyl") is intended to denote a group having from 6 to 15 carbons, consisting of an alkyl group that bears an aryl group. Examples of aralkyl groups include benzyl, phenethyl, benzhydryl and naphthylmethyl groups.

The term "alkylaryl" (or "alkaryl") is intended to denote a group having from 6 to 15 carbons, consisting of an aryl group that bears an alkyl group. Examples of aralkyl groups include methylphenyl, ethylphenyl and methylnaphthyl groups.

The term "arylsulfonyl" denotes an aryl group attached through a sulfonyl group, for example phenylsulfonyl. The term "alkylsulfonyl" denotes an alkyl group attached through a sulfonyl group, for example methylsulfonyl.

The term "alkoxycarbonyl" denotes a group of formula  $-C(=O)-O-R$  where R is alkyl, alkenyl, or alkynyl, where the alkyl, alkenyl, or alkynyl portions thereof can be optionally substituted as described herein.

The term "aryloxcarbonyl" denotes a group of formula  $-C(=O)-O-R$  where R is aryl, where the aryl portion thereof can be optionally substituted as described herein.

The terms "arylalkyloxy" or "aralkyloxy" are equivalent, and denote a group of formula  $-O-R'-R''$ , where  $R'$  is R is alkyl, alkenyl, or alkynyl which can be optionally substituted as described herein, and wherein  $R''$  denotes a aryl or substituted aryl group.

The terms "alkylarylloxy" or "alkarylloxy" are equivalent, and denote a group of formula  $-O-R'-R''$ , where  $R'$  is an aryl or substituted aryl group, and  $R''$  is alkyl, alkenyl, or alkynyl which can be optionally substituted as described herein.

As used herein, the term "aldehyde group" denotes a group that bears a moiety of formula  $-C(=O)-H$ . The term "ketone" denotes a moiety containing a group of formula  $-R-C(=O)-R=$ ,



where R and R= are independently alkyl, alkenyl, alkynyl, aryl, heteroaryl, aralkyl, or alkaryl, each of which may be substituted as described herein.

As used herein, the term "ester" denotes a moiety having a group of formula  $-R-C(=O)-O-R=$  or  $-R-O-C(=O)-R=$  where R and R= are independently alkyl, alkenyl, alkynyl, aryl, heteroaryl, aralkyl, or alkaryl, each of which may be substituted as described herein.

The term "ether" denotes a moiety having a group of formula  $-R-O-R=$  or where R and R= are independently alkyl, alkenyl, alkynyl, aryl, heteroaryl, aralkyl, or alkaryl, each of which may be substituted as described herein.

The term "crown ether" has its usual meaning of a cyclic ether containing several oxygen atoms. As used herein the term "organosulfur compound" denotes aliphatic or aromatic sulfur containing compounds, for example thiols and disulfides. The term "organometallic group" denotes an organic molecule containing at least one metal atom.

The term "organosilicon compound" denotes aliphatic or aromatic silicon containing compounds, for example alkyl and aryl silanes.

The term "carboxylic acid" denotes a moiety having a carboxyl group, other than an amino acid.

As used herein, the term "amino acid" denotes a molecule containing both an amino group and a carboxyl group. In some preferred embodiments, the amino acids are  $\alpha$ -,  $\beta$ -,  $\gamma$ - or  $\delta$ -amino acids, including their stereoisomers and racemates. As used herein the term "L-amino acid" denotes an  $\alpha$ -amino acid having the L configuration around the  $\alpha$ -carbon, that is, a carboxylic acid of general formula  $CH(COOH)(NH_2)$ -(side chain), having the L-configuration.

The term "D-amino acid" similarly denotes a carboxylic acid of general formula  $CH(COOH)(NH_2)$ -(side chain), having the D-configuration around the  $\alpha$ -carbon. Side chains of L-amino acids include naturally occurring and non-naturally occurring moieties. Non-naturally occurring (i.e., unnatural) amino acid side chains are moieties that are used in place of naturally occurring amino acid side chains in, for example, amino acid analogs. See, for example, Lehninger, *Biochemistry*, Second Edition, Worth Publishers, Inc, 1975, pages 72-77, incorporated herein by reference. Amino acid substituents may be attached through their carbonyl groups through the oxygen or carbonyl carbon thereof, or through their amino groups, or through functionalities residing on their side chain portions.

As used herein "polynucleotide" refers to a nucleic acid molecule and includes genomic DNA cDNA, RNA, mRNA and the like.

As used herein "antisense oligonucleotide" refers to a nucleic acid molecule that is complementary to at least a portion of a target nucleotide sequence of interest and specifically hybridizes to the target nucleotide sequence under physiological conditions.

As used herein "inhibitor of mismatch repair" refers to an agent that interferes with at least one function of the mismatch repair system of a cell and thereby renders the cell more susceptible to mutation.

As used herein "hypermutable" refers to a state in which a cell *in vitro* or *in vivo* is made more susceptible to mutation through a loss or impairment of the mismatch repair system.

As used herein "agents," "chemicals," and "inhibitors" when used in connection with inhibition of MMR refers to chemicals, oligonucleotides, analogs of natural substrates, and the like that interfere with normal function of MMR.

Methods for developing hypermutable cells and whole organisms have been discovered by taking advantage of the conserved mismatch repair (MMR) process of a host. Dominant negative alleles of MMR genes, when introduced into cells or transgenic animals, increase the rate of spontaneous mutations by reducing the effectiveness of DNA repair and thereby render the cells or animals hypermutable. Hypermutable microbes, protozoans, insects, mammalian cells, plants or whole animals can then be utilized to develop new mutations in a gene of interest. It has been discovered that chemicals that block MMR, and thereby render cells hypermutable, is an efficient way to introduce mutations in cells and genes of interest. In addition to destabilizing the genome of cells exposed to chemicals that inhibit MMR activity may be done transiently, allowing cells to become hypermutable, and removing the chemical exposure after the desired effect (e.g., a mutation in a gene of interest) is achieved. The chemicals that inhibit MMR activity that are suitable for use in the invention include, but are not limited to, anthracene derivatives, nonhydrolyzable ATP analogs, ATPase inhibitors, antisense oligonucleotides that specifically anneal to polynucleotides encoding mismatch repair proteins, DNA polymerase inhibitors, and exonuclease inhibitors. These chemicals can enhance the rate of mutation due to inactivation

of MMR yielding clones or subtypes with altered biochemical properties. Methods for identifying chemical compounds that inhibit MMR *in vivo* are also described herein.

The process of MMR, also called mismatch proofreading, is carried out by a group of protein complexes in cells ranging from bacteria to man (Harfe B.D., and S. Jinks-Robertson (2000) *An. Rev. Genet.* 34:359-399; Modrich, P. (1994) *Science* 266:1959-1960). An MMR gene is a gene that encodes for one of the proteins of such a mismatch repair complex. Although not wanting to be bound by any particular theory of mechanism of action, an MMR complex is believed to detect distortions of the DNA helix resulting from non-complementary pairing of nucleotide bases. The non-complementary base on the newer DNA strand is excised, and the excised base is replaced with the appropriate base, which is complementary to the older DNA strand. In this way, cells eliminate many mutations that occur as a result of mistakes in DNA replication.

Dominant negative alleles cause an MMR defective phenotype even in the presence of a wild-type allele in the same cell. An example of a dominant negative allele of an MMR gene is the human gene *hPMS2-134* (SEQ ID NO:25), which carries a truncating mutation at codon 134 (Nicolaidis, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641). The mutation causes the product of this gene to abnormally terminate at the position of the 134th amino acid, resulting in a shortened polypeptide containing the N-terminal 133 amino acids (SEQ ID NO:24). Such a mutation causes an increase in the rate of mutations, which accumulate in cells after DNA replication. Expression of a dominant negative allele of a mismatch repair gene results in impairment of mismatch repair activity, even in the presence of the wild-type allele.

The MMR process has been shown to be blocked by the use of nonhydrolyzable forms of ATP (Galio, L. *et al.* (1999) *Nucl. Acids Res.* 27:2325-2331; Allen, D.J. *et al.* (1997) *EMBO J.* 16:4467-4476; Bjornson, K.P. *et al.* (2000) *Biochem.* 39:3176-3183). However, it has not been demonstrated that chemicals can block MMR activity in cells. Such chemicals can be identified by screening cells for defective MMR activity. Cells from bacteria, yeast, fungi, insects, plants, animals, and humans can be screened for defective mismatch repair. Genomic DNA, cDNA, or mRNA from any cell can be analyzed for variations from the wild type sequences in cells or organisms grown in the presence of MMR blocking compounds.

Various techniques of screening can be used. The suitability of such screening assays, whether natural or artificial, for use in identifying hypermutable cells, insects, fungi, plants or animals can be evaluated by testing the mismatch repair activity caused by a compound or a mixture of compounds, to determine if it is an MMR inhibitor.

A cell, a microbe, or a whole organism such as an insect, fungus, plant or animal in which a chemical inhibitor of mismatch repair has been treated will become hypermutable. This means that the spontaneous mutation rate of such cells or whole organism is elevated compared to cells or animals without such treatment. The degree of elevation of the spontaneous mutation rate can be at least 2-fold, 5-fold, 10-fold, 20-fold, 50-fold, 100-fold, 200-fold, 500-fold, or 1000-fold that of the normal cell or animal. The use of chemical mutagens such as, but limited to, N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), methane sulfonate, dimethyl sulfonate, O6-methyl benzadine, ethyl methanesulfonate (EMS), methylnitrosourea (MNU), ethylnitrosourea (ENU), *etc.* can be used in MMR defective cells or whole organisms to increase the rates an additional 10 to 100 fold that of the MMR deficiency itself.

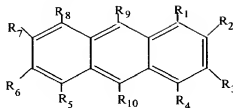
According to one aspect of the invention, a screening assay for identifying chemical inhibitors of MMR is developed and employed. A chemical compound can be in any form or class ranging from but not limited to amino acid, steroidal, aromatic, or lipid precursors. The chemical compound can be naturally occurring or made in the laboratory. The screening assay can be natural such as looking for altered endogenous repeats within an host organism's genome (as demonstrated in Figs. 4 and 5), or made in the laboratory using an MMR-sensitive reporter gene as demonstrated in Figs. 1-3).

The chemical compound can be introduced into the cell by supplementing the growth medium, or by intracellular delivery such as but not limited to using microinjection or carrier compounds.

According to another aspect of the invention, a chemical compound from the anthracene class can be exposed to MMR proficient cells or whole organism hosts, the host is grown and screened for subtypes containing genetically altered genes with new biochemical features.

The anthracene compounds that are suitable for use in the invention include, but are

not limited to anthracenes having the formula:



wherein R<sub>1</sub>-R<sub>10</sub> are independently hydrogen, hydroxyl, amino, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkynyl, substituted alkynyl, O-alkyl, S-alkyl, N-alkyl, O-alkenyl, S-alkenyl, N-alkenyl, O-alkynyl, S-alkynyl, N-alkynyl, aryl, substituted aryl, aryloxy, substituted aryloxy, heteroaryl, substituted heteroaryl, aralkyloxy, arylalkyl, alkylaryl, alkylaryloxy, arylsulfonyl, alkylsulfonyl, alkoxycarbonyl, aryloxy carbonyl, guanidino, carboxy, an alcohol, an amino acid, sulfonate, alkyl sulfonate, CN, NO<sub>2</sub>, an aldehyde group, an ester, an ether, a crown ether, a ketone, an organosulfur compound, an organometallic group, a carboxylic acid, an organosilicon or a carbohydrate that optionally contains one or more alkylated hydroxyl groups;

wherein said heteroalkyl, heteroaryl, and substituted heteroaryl contain at least one heteroatom that is oxygen, sulfur, a metal atom, phosphorus, silicon or nitrogen; and

wherein said substituents of said substituted alkyl, substituted alkenyl, substituted alkynyl,

substituted aryl, and substituted heteroaryl are halogen, CN, NO<sub>2</sub>, lower alkyl, aryl, heteroaryl, aralkyl, aralkyloxy, guanidino, alkoxycarbonyl, alkoxy, hydroxy, carboxy and amino;

and wherein said amino groups optionally substituted with an acyl group, or 1 to 3 aryl or lower alkyl groups;

or wherein any two of R<sub>1</sub>-R<sub>10</sub> can together form a polyether;

or wherein any two of R<sub>1</sub>-R<sub>10</sub> can, together with the intervening carbon atoms of the anthracene core, form a crown ether.

The method of the invention also encompasses inhibiting MMR with an anthracene of the above formula wherein R<sub>5</sub> and R<sub>6</sub> are hydrogen, and the remaining substituents are as described above.

The some embodiments, in the anthracene compound R<sub>1</sub>-R<sub>10</sub> are independently hydrogen, hydroxyl, alkyl, aryl, arylalkyl, or hydroxyalkyl. In other embodiments, R<sub>1</sub>-R<sub>10</sub> are independently

hydrogen, hydroxyl, methyl, ethyl, propyl, isopropyl, butyl, isobutyl, phenyl, tolyl, hydroxymethyl, hydroxypropyl, or hydroxybutyl.

In specific embodiments of the invention the anthracenes include, but are not limited to 1,2-dimethylantracene, 9,10-dimethyl anthracene, 7,8-dimethylantracene, 9,10-diphenylantracene, 9,10-dihydroxymethylantracene, 9-hydroxymethyl-10-methylantracene, dimethylantracene-1,2-diol, 9-hydroxymethyl-10-methylantracene-1,2-diol, 9-hydroxymethyl-10-methylantracene-3,4-diol, 9, 10-di-m-tolyantracene, and the like.

The chiral position of the side chains of the anthracenes is not particularly limited and may be any chiral position and any chiral analog. The anthracenes may also comprise a stereoisomeric forms of the anthracenes and includes any isomeric analog.

Examples of hosts are but not limited to cells or whole organisms from human, primate, mammal, rodent, plant, fish, reptiles, amphibians, insects, fungi, yeast or microbes of prokaryotic origin.

Yet another aspect of the invention is the use of ATP analogs capable of blocking ATPase activity required for MMR. MMR reporter cells are screened with ATP compound libraries to identify those compounds capable of blocking MMR *in vivo*. Examples of ATP analogs that are useful in blocking MMR activity include, but are not limited to, nonhydrolyzable forms of ATP such as AMP-PNP and ATP[gamma]S block the MMR activity (Galio, L. *et al.* (1999) *Nucl. Acids Res.* 27:2325-2331; Allen, D.J. *et al.* (1997) *EMBO J.* 16:4467-4476; Bjornson K.P. *et al.* (2000) *Biochem.* 39:3176-3183).

Yet another aspect of the invention is the use of nuclease inhibitors that are able to block the exonuclease activity of the MMR biochemical pathway. MMR reporter cells are screened with nuclease inhibitor compound libraries to identify compounds capable of blocking MMR *in vivo*. Examples of nuclease inhibitors that are useful in blocking MMR activity include, but are not limited to analogs of N-Ethylmaleimide, an endonuclease inhibitor (Huang, Y.C., *et al.* (1995) *Arch. Biochem. Biophys.* 316:485), heterodimeric adenine-chain-acridine compounds, exonuclease III inhibitors (Belmont P, *et al.*, *Bioorg Med Chem Lett* (2000) 10:293-295), as well as antibiotic compounds such as Heliquinomycin, which have helicase inhibitory activity (Chino, M, *et al.* *J. Antibiot. (Tokyo)* (1998) 51:480-486).

Another aspect of the invention is the use of DNA polymerase inhibitors that are able to block the polymerization required for mismatch-mediated repair. MMR reporter cells are screened with DNA polymerase inhibitor compound libraries to identify those compounds capable of blocking MMR *in vivo*. Examples of DNA polymerase inhibitors that are useful in blocking MMR activity include, but are not limited to, analogs of actinomycin D (Martin, S.J., et.al. (1990) *J. Immunol.* 145:1859), Aphidicolin (Kuwakado, K. et.al. (1993) *Biochem. Pharmacol.* 46:1909) 1-(2'-Deoxy-2'-fluoro-beta-L-arabinofuranosyl)-5-methyluracil (L-FMAU) (Kukhanova M, et.al., *Biochem Pharmacol* (1998) 55:1181-1187), and 2',3'-dideoxyribonucleoside 5'-triphosphates (ddNTPs) (Ono, K., et.al., *Biomed Pharmacother* (1984) 38:382-389).

In yet another aspect of the invention, antisense oligonucleotides are administered to cells to disrupt at least one function of the mismatch repair process. The antisense polynucleotides hybridize to MMR polynucleotides. Both full-length and antisense polynucleotide fragments are suitable for use. "Antisense polynucleotide fragments" of the invention include, but are not limited to polynucleotides that specifically hybridize to an MMR encoding RNA (as determined by sequence comparison of nucleotides encoding the MMR to nucleotides encoding other known molecules). Identification of sequences that are substantially unique to MMR-encoding polynucleotides can be ascertained by analysis of any publicly available sequence database and/or with any commercially available sequence comparison programs. Antisense molecules may be generated by any means including, but not limited to chemical synthesis, expression in an *in vitro* transcription reaction, through expression in a transformed cell comprising a vector that may be transcribed to produce antisense molecules, through restriction digestion and isolation, through the polymerase chain reaction, and the like.

Antisense oligonucleotides, or fragments thereof may include the nucleotide sequences set forth in SEQ ID NOs:15, 17, 19, 21, 23, 25, 27, and 29 or sequences complementary or homologous thereto, for example. Those of skill in the art recognize that the invention may be predicted using any MMR gene. Specifically, antisense nucleic acid molecules comprise a sequence complementary to at least about 10, 15, 25, 50, 100, 250 or 500 nucleotides or an entire MMR encoding sequence. Preferably, the antisense

oligonucleotides comprise a sequence complementary to about 15 consecutive nucleotides of the coding strand of the MMR encoding sequence.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an MMR protein. The coding strand may also include regulatory regions of the MMR sequence. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the protein coding region of human PMS2 corresponds to the coding region SEQ ID NO:17). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding an MMR protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions (UTR)).

Preferably, antisense oligonucleotides are directed to regulatory regions of a nucleotide sequence encoding an MMR protein, or mRNA corresponding thereto, including, but not limited to, the initiation codon, TATA box, enhancer sequences, and the like. Given the coding strand sequences provided herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of an MMR mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of an MMR mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of an MMR mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length.

Screening is any process whereby a chemical compound is exposed to a cell or whole organism. The process of screening can be carried out using but not limited to a whole animal, plant, insect, microbe, or by using a suspension of one or more isolated cells in culture. The cell can be any type of eukaryotic or prokaryotic cell, including, for example, cells isolated from humans or other primates, mammals or other vertebrates, invertebrates, and single celled organisms such as protozoa, yeast, or bacteria.



In general, screening will be carried out using a suspension of cells, or a single cell, but other methods can also be applied as long as a sufficient fraction of the treated cells or tissue is exposed so that isolated cells can be grown and utilized. Techniques for chemical screening are well known to those in the art. Available techniques for screening include cell-based assays, molecular assays, and whole organism-based assays. Compounds can be added to the screening assays of the invention in order to identify those agents that are capable of blocking MMR in cells.

The screening assays of the invention provide a system wherein a cell, cells or a whole organism is contacted with a candidate compound and then tested to determine whether mismatch repair has been adversely affected. The method in which MMR is analyzed may be any known method, including, but not limited to analysis of the molecular sequence of the MMR gene, and analyzing endogenous repeats in the subject's genome. Further, the invention provides a convenient assay to analyze the effects of candidate agents on reporter genes transfected into cells.

MMR-inhibitors identified by the methods of the invention can be used to generate new mutations in one or more gene(s) of interest. A gene of interest can be any gene naturally possessed by a cell line, microbe or whole organism. An advantage of using chemicals rather than recombinant technologies to block MMR are that the process is faster; there is no need to produce stable clones with a knocked out MMR gene or a clone expressing a dominant negative MMR gene allele. Another advantage is that host organisms need not be screened for integrated knock out targeting vectors or stable expression of a dominant negative MMR gene allele. Finally, once a cell, plant or animal has been exposed to the MMR-blocking compound and a new output trait is generated, the MMR process can be restored by removal of compound. Mutations can be detected by analyzing the genotype of the cell, or whole organism, for example, by examining the sequence of genomic DNA, cDNA, messenger RNA, or amino acids associated with the gene of interest. Mutations can also be detected by screening for new output traits such as hypoxanthine-guanine phosphoribosyltransferase (HPRT) revertants. A mutant polypeptide can be detected by identifying alterations in electrophoretic mobility, spectroscopic properties, or other physical or structural characteristics of a protein encoded by a mutant gene. One can also screen for

altered function of the protein *in situ*, in isolated form, or in model systems. One can screen for alteration of any property of the cell, plant or animal associated with the function of the gene of interest.

Several advantages exist in generating genetic mutations by blocking MMR *in vivo* in contrast to general DNA damaging agents such as MNNG, MNU and EMS. Cells with MMR deficiency have a wide range of mutations dispersed throughout their entire genome in contrast to DNA damaging agents such as MNNG, MNU, EMS and ionizing radiation. Another advantage is that mutant cells that arise from MMR deficiency are diploid in nature and do not lose large segments of chromosomes as is the case of DNA damaging agents such as EMS, MNU, and ionizing radiation (Honma, M. *et al.* (1997) *Mutat. Res.* 374:89-98). This unique feature allows for subtle changes throughout a host's genome that leads to subtle genetic changes yielding genetically stable hosts with commercially important output traits.

The invention also encompasses blocking MMR *in vivo* and *in vitro* and further exposing the cells or organisms to a chemical mutagen in order to increase the incidence of genetic mutation.

The invention also encompasses withdrawing exposure to inhibitors of mismatch repair once a desired mutant genotype or phenotype is generated such that the mutations are thereafter maintained in a stable genome.

The above disclosure generally describes the present invention. A more complete understanding can be obtained by reference to the following specific examples, which are provided herein for purposes of illustration only, and are not intended to limit the scope of the invention.

### EXAMPLES

#### **EXAMPLE 1: Generation of a cell-based screening assay to identify chemicals capable of inactivating mismatch repair *in vivo*.**

A hallmark of MMR deficiency is the generation of unstable microsatellite repeats in the genome of host cells (Peinado, M.A. *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:10065-10069; Strand, M. *et al.* (1993) *Nature* 365:274-276; Parsons, R. *et al.* (1993) *Cell* 75:1227-1236). This phenotype is referred to as microsatellite instability (MI) (Harfe, B.D.

and S. Jinks-Robertson (2000) *Ann. Rev. Genet.* 34:359-399; Modrich, P. (1994) *Science* 266:1959-1960; Peinado, M.A. *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:10065-10069; Perucho, M. (1996) *Biol. Chem.* 377:675-684; Hoang, J.M. *et al.* (1997) *Cancer Res.* 57:300-303; Strand, M. *et al.* (1993) *Nature* 365:274-276). MI consists of deletions and/or insertions within repetitive mono-, di- and/or tri nucleotide repetitive sequences throughout the entire genome of a host cell. Extensive genetic analysis of eukaryotic cells have found that the only biochemical defect that is capable of producing MI is defective MMR (Harfe, B.D. and S. Jinks-Robertson (2000) *Ann. Rev. Genet.* 34:359-399; Modrich, P. (1994) *Science* 266:1959-1960; Peinado, M.A. *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:10065-10069; Perucho, M. (1996) *Biol. Chem.* 377:675-684; Hoang, J.M. *et al.* (1997) *Cancer Res.* 57:300-303; Strand, M. *et al.* (1993) *Nature* 365:274-276). In light of this unique feature that defective MMR has on promoting microsatellite instability, endogenous MI is now used as a biochemical marker to survey for lack of MMR activity within host cells (Hoang, J.M. *et al.* (1997) *Cancer Res.* 57:300-303).

A method used to detect MMR deficiency in eukaryotic cells is to employ a reporter gene that has a polynucleotide repeat inserted within the coding region that disrupts its reading frame due to a frame shift. In the case where MMR is defective, the reporter gene will acquire random mutations (*i.e.*, insertions and/or deletions) within the polynucleotide repeat yielding clones that contain a reporter with an open reading frame. This reporter gene can be of any biochemical pathway such as but not limited to  $\beta$ -glucuronidase,  $\beta$ -galactosidase, neomycin resistant gene, hygromycin resistance gene, green fluorescent protein, and the like. A schematic diagram of MMR-sensitive reporters are shown in Fig. 1, where the polynucleotide repeat can consist of mono-, di-, tri- or tetra-nucleotides. We have employed the use of a  $\beta$ -galactosidase MMR-sensitive reporter gene to measure for MMR activity in H36 cells, which are a murine hybridoma cell line. The reporter construct used is called pCAR-OF, which contains a hygromycin resistance (HYG) gene plus a  $\beta$ -galactosidase gene with a 29 bp out-of-frame poly-CA tract inserted at the 5' end of its coding region. The pCAR-OF reporter cannot generate  $\beta$ -galactosidase activity unless a frame-restoring mutation (*i.e.*, insertion or deletion) arises following transfection. This line has been shown to be sensitive to inactivated MMR where using a dominant negative MMR gene allele has found

this condition to result in the production of  $\beta$ -galactosidase (unpublished data). An example of these data using the dominant negative PMS134 allele is shown in Table 1. Briefly, H36 cells were each transfected with an expression vector containing the PMS134 allele (referred to as HB134) or empty vector and the pCAR-OF vector in duplicate reactions using the protocol below. The PMS134 gene is cloned into the pEF expression vector, which contains the elongation factor promoter upstream of the cloning site followed by a mammalian polyadenylation signal. This vector also contains the NEO<sup>r</sup> gene that allows for selection of cells in G418 to identify those retaining this plasmid. Briefly, cells were transfected with 1  $\mu$ g of the PMS134 or empty vector using polyliposomes following the manufacturer's protocol (Life Technologies). Cells were then selected in 0.5 mg/ml of G418 for 10 days and G418 resistant cells were pooled together to analyze for gene expression. PMS134 positive cells, which were determined by RT-PCR and western blot (not shown) were expanded and transfected with the pCAR-OF reporter gene that contains a hygromycin (HYG) resistance gene as reporter using the protocol described above. Cells were selected in 0.5 mg/ml G418 and 0.5mg/ml HYG to select for cells retaining both the MMR effector and the pCAR-OF reporter plasmids. All cultures transfected with the pCAR vector resulted in a similar number of HYG/G418 resistant cells. Cultures were then expanded and tested for  $\beta$ -galactosidase activity *in situ* as well as by biochemical analysis of cell extracts. For *in situ* analysis, 100,000 cells were harvested and fixed in 1% glutaraldehyde, washed in phosphate buffered saline solution and incubated in 1 ml of X-gal substrate solution [0.15 M NaCl, 1 mM MgCl<sub>2</sub>, 3.3 mM K<sub>4</sub>Fe(CN)<sub>6</sub>, 3.3 mM K<sub>3</sub>Fe(CN)<sub>6</sub>, 0.2% X-Gal ] in 24 well plates for 2 hours at 37°C. Reactions were stopped in 500 mM sodium bicarbonate solution and transferred to microscope slides for analysis. Three fields of 200 cells each were counted for blue ( $\beta$ -galactosidase positive cells) or white ( $\beta$ -galactosidase negative cells) to assess for MMR inactivation. Table 1 shows the results from these studies. While no  $\beta$ -galactosidase positive cells were observed in H36 empty vector cells and 10% of the cells per field were  $\beta$ -galactosidase positive in HB134 cultures.

**Table 1.**  $\beta$ -galactosidase expression of H36 empty vector and HB134 cells transfected with pCAR-OF reporter vectors. Cells were transfected with the pCAR-OF reporter plasmid.

Transfected cells were selected in HYG and G418, expanded and stained with X-gal solution to measure for  $\beta$ -galactosidase activity (blue colored cells). 3 fields of 200 cells each were analyzed by microscopy. The results below represent the mean +/- standard deviation of these experiments.

Table 1.

CELL LINE	# BLUE CELLS
H36 empty vector	0 +/- 0
HB134	20 +/- 3

Cultures can be further analyzed by biochemical assays using cell extracts to measure  $\beta$ -galactosidase activity as previously described (Nicolaidis, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641).

The data described in Table 1 show that by inhibiting the MMR activity of an MMR proficient cell host can result in MI and the altering of microsatellites in the pCAR-OF vector results in cells that produce functional  $\beta$ -galactosidase enzyme. The use of the H36pCAR-OF cell line can now be used to screen for chemicals that are able to block MMR of the H36 cell line.

#### EXAMPLE 2: Screening assays for identifying chemical blockers of MMR.

A method for screening chemical libraries is provided in this example using the H36pCAR-OF cell line described in Example 1. This cell line is a hardy, stable line that can be formatted into 96-well microtiter plates for automated screening for chemicals that specifically block MMR. An overview of the screening process is given in Figure 2, however, the process is not limited to the specifications within this example. Briefly, 10,000 cells in a total volume of 0.1ml of growth medium (RPMI1640 plus 10% fetal bovine serum) are added to 96-well microtiter plates containing any variety of chemical compounds. Cells are grown for 14-17 days at 37°C in 5%CO<sub>2</sub>. Cells are then lysed in the growth medium with 50ul of lysis buffer containing 0.1 M Tris buffer (pH 8.0), 0.1% Triton X-100, 45 mM 2-mercaptoethanol, 1mM MgCl<sub>2</sub>, 0.1 M NaPO<sub>4</sub> and 0.6 mg/ml Chlorophenol-red-  $\beta$ -

D-galactopyranoside (CPRG, Roche). Reactions are incubated for 1 hour, terminated by the addition of 50  $\mu$ ls of 0.5 M  $\text{Na}_2\text{CO}_3$ , and analyzed by spectrophotometry at 576 nm.

Experimental wells are compared to untreated or vehicle treated wells to identify those with increased  $\beta$ -galactosidase activity. Compounds producing MMR blocking activity are then further analyzed using different cell lines containing the pCAR-OF plasmid to measure the ability to block MMR as determined by MI in MMR proficient hosts by analyzing endogenous microsatellites for instability using assays described below.

### EXAMPLE 3: Defining MMR blocking chemicals.

The identification of chemical inhibitors of MMR can be difficult in determining those that are standard mutagens from those that induce genomic instability via the blockade of MMR. This Example teaches of a method for determining blockers of MMR from more general mutagens. Once a compound has been identified in the assay described above, one can determine if the compound is a general mutagen or a specific MMR blocker by monitoring mutation rates in MMR proficient cells and a controlled subclone that is MMR defective. One feature of MMR deficiency is the increased resistance to toxicity of DNA alkylating agents that allows for enhanced rates of mutations upon mutagen exposure (Liu, L., et.al. *Cancer Res* (1996) 56:5375-5379). This unique feature allows for the use of a MMR proficient cell and a controlled line to measure for enhanced activity of a chemical compound to induce mutations in MMR proficient vs MMR deficient lines. If the compound is a true inhibitor of MMR then genetic mutations should occur in MMR proficient cells while no "enhanced" mutation rate will be found in already MMR defective cells. Using these criteria chemicals such as ICR191, which induces frameshift mutations in mammalian cells would not be considered a MMR blocking compound because of its ability to produce enhanced mutation rates in already MMR defective cell lines (Chen, W.D., et.al. *J Natl Cancer Inst.* (2000) 92:480-485). These screening lines include the but are not limited those in which a dominant negative MMR gene has been introduced such as that described in EXAMPLE 1 or those in which naturally MMR deficient cells such as HCT116 has been cured by introduction of a complementing MMR gene as described (Chen, W.D., et.al. *J Natl Cancer Inst.* (2000) 92:480-485).

**EXAMPLE 4: Identification of chemical inhibitors of MMR *in vivo*.**

MMR is a conserved post replicative DNA repair mechanism that repairs point mutations and insertion/deletions in repetitive sequences after cell division. The MMR requires an ATPase activity for initiation complex recognition and DNA translocation. *In vitro* assays have shown that the use of nonhydrolyzable forms of ATP such as AMP-PNP and ATP[gamma]S block the MMR activity (Galio, L. *et al.* (1999) *Nucl. Acids Res.* 27:2325-2331; Allen, D.J. *et al.* (1997) *EMBO J.* 16:4467-4476; Bjornson K.P. *et al.* (2000) *Biochem.* 39:3176-3183).

The use of chemicals to inhibit endogenous MMR *in vivo* has not been distinguished in the public domain. In an attempt to identify chemicals that can inhibit MMR *in vivo*, we used our H36pCAR-OF screening assay to screen for chemicals that are able to cause microsatellite instability and restoration of  $\beta$ -galactosidase activity from the pCAR-OF vector, an effect that can only be caused due to MMR deficiency. In our screening assays we used a variety of classes of compounds ranging from steroids such as pontasterone to potent alkylating agents such as EMS, to kinase and other enzyme inhibitors. Screens identified one class of chemicals that were capable of generating  $\beta$ -galactosidase positive cells. These molecules were derived from the anthracene class. An example of one such anthracene derivative for the purposes of this application is a molecule called 9,10-dimethylanthracene, referred to from here on as DMA. Fig. 3 shows the effect of DMA in shifting the pCAR-OF reporter plasmid. In contrast, general DNA alkylating agents such as EMS or MNNG did not result in MI and/or the shifting of the polynucleotide tract in the pCAR-OF reporter.

The most likely explanation for the differences in  $\beta$ -galactosidase activity was that the DMA compound disturbed MMR activity, resulting in a higher frequency of mutation within the pCAR-OF reporter and re-establishing the ORF. To directly test the hypothesis that MMR was altered, we employ a biochemical assay for MMR with the individual clones as described by Nicolaides *et al.*, 1997 (Nicolaides, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641). Nuclear extracts are prepared from the clones and incubated with heteroduplex substrates containing either a /CA\ insertion-deletion or a G/T mismatch under conditions described previously. The /CA\ and G/T heteroduplexes are used to test repair from the 3'

and 5' directions, respectively as described (Nicolaides, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641).

### **Biochemical assays for mismatch repair.**

#### **Enzymatic Repair Assays:**

MMR activity in nuclear extracts is performed as described, using 24 fmol of substrate (Nicolaides, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641). Complementation assays are done by adding ~ 100 ng of purified MutLa or MutSa components to 100 µg of nuclear extract, adjusting the final KCl concentration to 100 mM (Nicolaides, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641). The substrates used in these experiments contain a strand break 181 nucleotides 5' or 125 nucleotides 3' to the mismatch.

#### **Biochemical Activity Assays:**

To demonstrate the direct effect to small molecules on MMR proteins, molecular assays such as mismatch binding and MMR complex formation are performed in the presence or absence of drug. Briefly, MMR gene cDNAs are PCR amplified using primers encompassing the entire coding regions of the known MMR proteins MSH2 (SEQ ID NO:20), GTBP (SEQ ID NO:26), MLH1 (SEQ ID NO:22), human PMS2 (SEQ ID NO:16), mouse PMS2 (SEQ ID NO:14), PMS1 (SEQ ID NO:18), and MHS3 (SEQ ID NO:28) from any species with a sense primer containing a T7 promoter and a Kozak translation signal as previously described (Nicolaides, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641). The coding regions of known MMR proteins include the sequences shown in Table 3 for mouse *PMS2* (SEQ ID NO:15), human *PMS2* (SEQ ID NO:17), human *PMS1* (SEQ ID NO:19), human *MSH2* (SEQ ID NO:21), human *MLH1* (SEQ ID NO:23), and human *MSH3* (SEQ ID NO:29). Products are transcribed and translated using the TNT system (Promega). An example of PCR primers and *in vitro* transcription-translation reactions are listed below.



**In vitro transcription-translation:**

Linear DNA fragments containing *hPMS2* (SEQ ID NO:17) and *hMLH1* (SEQ ID NO:23) cDNA sequences were prepared by PCR, incorporating sequences for *in vitro* transcription and translation in the sense primer. A full-length *hMLH1* fragment was prepared using the sense primer 5'-ggatcctaatacgaactcactatagggagaccaccatgtcgttcgtggcaggg-3' (SEQ ID NO:1)(codons 1-6) and the antisense primer 5'-taagtcttaagtgtaccaac-3' (SEQ ID NO:2)(located in the 3' untranslated region, nt 2411-2433), using a wild-type *hMLH1* cDNA clone as template. A full-length *hPMS2* fragment was prepared with the sense primer 5'-ggatcctaatacgaactcactatagggagaccaccatggaacaattgctgcgg-3' (SEQ ID NO:3)(codons 1-6) and the antisense primer 5'-aggttagtgaagactctgtc-3' (SEQ ID NO:4)(located in 3' untranslated region, nt 2670-2690) using a cloned *hPMS2* cDNA as template. These fragments were used to produce proteins via the coupled transcription-translation system (Promega). The reactions were supplemented with <sup>35</sup>S-labelled methionine or unlabelled methionine. Lower molecular weight bands are presumed to be degradation products and/or polypeptides translated from alternative internal methionines.

To study the effects of MMR inhibitors, assays are used to measure the formation of MLH1 and PMS2 with or without compound using polypeptides produced in the TNT System (Promega) followed by immunoprecipitation (IP). To facilitate the IP, tags may be placed at the C-terminus of the PMS2 protein to use for antibody binding or antibodies directed to the MMR protein itself can be used for IP.

**Immunoprecipitations:**

Immunoprecipitations are performed on *in vitro* translated proteins by mixing the translation reactions with 1 µg of the MLH1 specific monoclonal antibody (mAb) MLH14 (Oncogene Science, Inc.), a polyclonal antibody generated to codons 2-20 of hPMS2 described above, or a polyclonal antibody generated to codons 843-862 of hPMS2 (Santa Cruz Biotechnology, Inc.) in 400 µl of EBC buffer (50 mM Tris, pH 7.5, 0.1 M NaCl, 0.5% NP40). After incubation for 1 hr at 4°C, protein A sepharose (Sigma) is added to a final concentration of 10% and reactions are incubated at 4°C for 1 hour. Proteins bound to protein

A are washed five times in EBC and separated by electrophoresis on 4-20% Tris-glycine gels, which are then dried and autoradiographed.

Compounds that block heterodimerization of mutS or mutL proteins can now be identified using this assay.

**EXAMPLE 5: Use of chemical MMR inhibitors yields microsatellite instability in human cells**

In order to demonstrate the global ability of a chemical inhibitor of MMR in host cells and organisms, we treated human HEK293 cells (referred to as 293 cells) with DMA and measured for microsatellite instability of endogenous loci using the BAT26 diagnostic marker (Hoang J.M. *et al.* (1997) *Cancer Res.* 57:300-303). Briefly,  $10^5$  cells were grown in control medium or 250  $\mu$ M DMA, a concentration that is found to be non-toxic, for 14 to 17 days. Cells are then harvested and genomic DNA isolated using the salting out method (Nicolaidis, N.C. *et al.* (1991) *Mol. Cell. Biol.* 11:6166-6176.).

Various amounts of test DNAs from HCT116 (a human colon epithelial cell line) and 293 were first used to determine the sensitivity of our microsatellite test. The BAT26 alleles are known to be heterogeneous between these two cell lines and the products migrate at different molecular weights (Nicolaidis personal observation). DNAs were diluted by limiting dilution to determine the level of sensitivity of the assay. DNAs were PCR amplified using the BAT26F: 5'-tgactacttttgacttcagcc-3' (SEQ ID NO:43) and the BAT26R: 5'-aaccattcaacatttttaaccc-3' (SEQ ID NO:44) primers in buffers as described (Nicolaidis, N.C. *et al.* (1995) *Genomics* 30:195-206). Briefly 1  $\mu$ g to 100 ngs of DNA were amplified using the following conditions: 94°C for 30 sec, 58°C for 30 sec, 72°C for 30 sec for 30 cycles. PCR reactions were electrophoresed on 12% polyacrylamide TBE gels (Novex) or 4% agarose gels and stained with ethidium bromide. These studies found that 0.1 ng of genomic DNA was the limit of detection using our conditions.

To measure for microsatellite stability in 293 cells grown with or without DMA, 0.1 ngs of DNA from DMA-treated or control 293 cells were amplified using the reaction conditions above. Forty individual reactions were carried out for each sample to measure for minor alleles. Fig. 4A shows a typical result from these studies whereby BAT26 alleles were

amplified from DMA-treated and untreated cells and analyzed on 12% PAGE gels (Novex). Alleles from DMA-treated cells showed the presence of an altered allele (asterisk) that migrated differently from the wild type allele. No altered alleles were found in the MMR-proficient control cells as expected since MI only occurs in MMR defective cell hosts. To confirm these data, PCRs were repeated using isolated BAT26 products. Primers and conditions were the same as described above except that reactions were amplified for 20 cycles. PCR products were gel-purified and cloned into T-tailed vectors (InVitrogen) as suggested by the manufacturer. Recombinant clones from DMA-treated and control cells were screened by PCR again using the BAT26 primers. Fifty bacterial colonies were analyzed for BAT26 structure by directly adding an aliquot of live bacteria to the PCR mix. PCR reactions were carried out as described above, and products were electrophoresed on 4% agarose gels and stained with ethidium bromide. As shown in Figure 4B, microsatellites from DMA-treated cells had alterations (asterisks) that made the marker length larger or smaller than the wild type allele found in control cells.

To confirm that these differences in molecular weight were due to shifts within the polynucleotide repeat, a hallmark of defective MMR, five clones from each sample were sequenced using an ABI automated sequencer with an M13-R primer located in the T-tail vector backbone. Sequence analysis revealed that the control cell clone used in our studies was homozygous for the BAT26 allele with a 26nt polyA repeat. Cells treated with DMA found multiple alleles ranging from the wild-type with 26 polyA repeat to shorter alleles (24 polyA repeat) and larger alleles (28 polyA repeat) (Fig. 5).

These data corroborate the H36pCAR data in Example 1 and Fig. 3 and demonstrates the ability to block MMR with a chemical in a range of hosts.

**Example 6: Chemical inhibitors of MMR generate DNA hypermutability in Plants and new phenotypes.**

To determine if chemical inhibitors of MMR work across a diverse array of organisms, we explored the activity of DMA on *Arabidopsis thaliana* (AT), a member of the mustard plant family, as a plant model system to study the effects of DMA on generating MMR deficiency, genome alterations, and new output traits.

Briefly, AT seeds were sterilized with straight commercial bleach and 100 seeds were plated in 100mm Murashige and Skoog (MS) phytagar (Life Technology) plates with increasing amounts of DMA (ranging from 100 $\mu$ M to 50mM). A similar amount of seeds were plated on MS phytagar only or in MS phytagar with increasing amounts of EMS (100 $\mu$ M to 50mM), a mutagen commonly used to mutate AT seeds (McCallum, C.M.*et al.* (2000) *Nat. Biotechnol.* 18:455-457). Plates were grown in a temperature-controlled, fluorescent-lighted humidifier (Percival Growth Chamber) for 10 days. After 10 days, seeds were counted to determine toxicity levels for each compound. Table 2 shows the number of healthy cells/treatment as determined by root formation and shoot formation. Plantlets that were identical to untreated seeds were scored healthy. Seeds with stunted root or shoot formation were scored intermediate (inter). Non-germinated seeds were scored dead.

**Table 2: Toxicity curve of DMA and EMS on *Arabidopsis* (per 100 cells)**

	0	0.1	0.5	1.0	2.5	5.0	10	12.5	25	50
<b>DMA</b>										
Healthy	100	94	99	99	85	85	65	0	0	0
Inter	0	0	0	0	20	15	32	85	100	0
Dead	0	0	0	0	0	0	0	0	0	100
<b>EMS</b>										
Healthy	99	100	45	25	0	0	0	0	0	0
Inter	0	0	54	75	0	0	0	0	0	0
Dead	0	0	0	0	100	100	100	100	100	87

The data in Table 2 show that DMA toxicity occurs at 10mM of continuous culture, while toxicity occurs at 250  $\mu$ M for EMS. Next, 50 seeds were plated in two 150mm dishes containing 2mM DMA, 250  $\mu$ M EMS or no drug. Seeds were grown for 10 days and then 10 plants from each plate were transferred to soil. All plants appeared to be similar in color and height. Plants were grown at room temperature with daily cycles of 18 hr light and 6 hr dark.

After 45 days seeds are harvested from siliques and stored in a desiccator at 4°C for 72 hours. Seeds are then sterilized and 100 seeds from each plant is sown directly into water-saturated soil and grown at room temperature with daily cycles of 18 hr light and 6 hr dark. At day 10 phenotypically distinct plants were found in 7 out of 118 DMA treated while no phenotypic difference was observed in 150 EMS-treated or 150 control plants. These 7 altered plants were light green in color and appeared to grow slower. Fig. 6 shows a typical difference between the DMA altered plant and controls. DMA-exposed plants produced offspring that were yellow in appearance in contrast to dark green, which is always found in wild-type plants. In addition, the yellow plants were also shorter. After 30 days, most wild-type plants produced flowers and siliques, while the 7 mutants just began flowering. After 45 days, control plants were harvested while mutant plants were harvested 10 to 15 days later. No such effects were observed in 150 plantlets from EMS treated plants.

The effect of DMA on MMR was confirmed by monitoring the structure of endogenous polynucleotide repeat markers within the plant genome. DNA was extracted using the DNAzol method following the manufacturer's protocol (Life Technology). Briefly, two leaves were harvested from DMA, EMS or untreated plants and DNA was extracted. DNAs were quantified by optical density using a BioRad Spectrophotometer. In *Arabidopsis*, a series of poly-A (A)<sub>n</sub>, (CA)<sub>n</sub> and (GA)<sub>n</sub> markers were found as a result of EMBL and GenBank database searches of DNA sequence data generated as a result of the *Arabidopsis* genome-sequencing project. Two markers that are naturally occurring, ATHACS and Nga128 are used to monitor microsatellite stability using primers described (Bell, C.J. and J.R. Ecker (1994) *Genomics* 19:137-144). ATHACS has a stretch of thirty-six adenine repeats (A)<sub>36</sub> whereas Nga128 is characterized by a di-nucleotide AG repeat that is repeated nineteen times (AG)<sub>19</sub>, while the Nga280 marker contains a polyAG repeat marker with 15 dinucleotides. DMA-mediated alterations of these markers are measured by a PCR assay.

Briefly, the genomic DNA is amplified with specific primers in PCR reaction buffers described above using 1-10ng plant genomic DNA. Primers for each marker are listed below:

nga280:

nga280-F: 5'-CTGATCTCACGGACAATAGTGC-3' (SEQ ID NO:5)

nga280-R: 5'-GGCTCCATAAAAAGTGCACC-3' (SEQ ID NO:6)

nga128:

nga128-F: 5'-GGTCTGTTGATGTCGTAAGTCG-3' (SEQ ID NO:7)

nga128-R: 5'-ATCTTGAACCTTTAGGGAGGG-3' (SEQ ID NO:8)

ATHACS:

ATHACS-F: 5'-AGAAGTTTAGACAGGTAC-3' (SEQ ID NO:9)

ATHACS-R: 5'-AAATGTGCAATTGCCTTC-3' (SEQ ID NO:10)

Cycling conditions are 94°C for 15 seconds, 55°C for 15 seconds and 72°C for 30 seconds, conditions that have been demonstrated to efficiently amplify these two markers (personal observation, Morphotek). PCR products are analyzed on 3.5% metaphor agarose gel in Tris-Acetate-EDTA buffer following staining with ethidium bromide.

Another method used to demonstrate that biochemical activity of a plant host's MMR is through the use of reporter genes disrupted by a polynucleotide repeat, similar to that described in Example 1 and Fig. 1. Due to the high endogenous  $\beta$ -galactosidase background, we engineered a plant compatible MMR-sensitive reporter gene consisting of the  $\beta$ -glucuronidase (GUS) gene with a mononucleotide repeat that was inserted just downstream of the initiation codon. Two reporter constructs were generated. pGUS-OF, contained a 20 base adenine repeat inserted just downstream of the initiating methionine that resulted in a frameshift, therefore producing a nonfunctional enzyme. The second, pGUS-IF, contained a 19 base adenine repeat that retained an open reading frame and served as a control for  $\beta$ -glucuronidase activity. Both constructs were generated by PCR using the pBI-121 vector (Life Technologies) as template. The antisense primer was directed to the 3' end of the Nopaline Synthase (NOS) polytermination sequence contained within the pBI-121 plasmid and contained a unique *EcoRI* restriction site to facilitate cloning of the vector into the pBI-121 binary vector backbone. The sense primers contained a unique *BamHI* restriction site to facilitate cloning of the chimeric GUS reporter gene into the pBI-121 binary vector backbone. The primers used to generate each reporter are:

1. sense primer for pGUS-IF (uidA-ATG-polyA-IF):

5'- CCC GGA TCC ATG TTA AAA AAA AAA AAA AAA AAA CGT CCT GTA GAA ACC-3' (SEQ ID NO:11)

2. sense primer for pGUS-OF (uidA-ATG-polyA-OF):

5'- CCC GGA TCC ATG TTA AAA AAA AAA AAA AAA AAA ACG TCC TGT AGA AAC C-3' (SEQ

ID NO:12)

## 3. antisense primer (Nos-term):

5'- CCC GAA TTC CCC GAT CTA GTA ACA TAG ATG-3' (SEQ ID NO:13)

PCR amplifications were carried out using reaction buffers described above.

Reactions were performed using 1 ng of pBI-121 vector as template (Life Technologies) and the appropriate corresponding primers. Amplifications were carried at 94°C for 30 seconds, 54°C for 60 seconds and 72°C for 60 seconds for 25 cycles. PCR products of the expected molecular weight was gel purified, cloned into T-tailed vectors (InVitrogen), and sequenced to ensure authentic sequence using the following primers: CaMV-FORW. [= 5'-gat atc tcc act gac gta ag-3'] (SEQ ID NO:30) for sequencing from the CaMV promoter into the 5' end of GUS cDNAs; NOSpA-42F [= 5'-tgt tgc cgg tct tgc gat g-3'] (SEQ ID NO:31) for sequencing of the NOS terminator; NOSpA-Cend-R [= 5'-ccc gat cta gta aca tag atg-3'] (SEQ ID NO:32) for sequencing from the NOS terminator into the 3' end of the GUS cDNAs; GUS-63F [= 5'-cag tct gga tgc cga aaa ctg-3'] (SEQ ID NO:33), GUS-441F [= 5'-ggt gat tac cga cga aaa cg-3'] (SEQ ID NO:34), GUS-825F [= 5'-agt gaa ggg cga aca gtt cc-3'] (SEQ ID NO:35), GUS-1224F [= 5'-gag tat tgc caa cga acc-3'] (SEQ ID NO:36), GUS-1596F [= 5'-gta tca ccg cgt ctt tga tc-3'] (SEQ ID NO:37), GUS-265R [= 5'-cga aac gca gca cga tac g-3'] (SEQ ID NO:38), GUS-646R [= 5'-gtt caa cgc tga cat cac c-3'] (SEQ ID NO:39), GUS-1033R [= 5'-cat gtt cat ctg ccc agt cg-3'] (SEQ ID NO:40), GUS-1425R [= 5'-gct ttg gac ata cca tcc-3'] (SEQ ID NO:41), and GUS-1783R [= 5'-cac cga agt tca tgc cag-3'] (SEQ ID NO:42) for the sequence of the full length GUS cDNAs. No mutation were found in either the OF or IF version of the GUS cDNA, and the expected frames for both cDNAs were also confirmed. pCR-IF-GUS and pCR-OF-GUS plasmids were subsequently digested with the BamH I and EcoR I restriction endonucleases, to generate DNA fragments containing the GUS cDNA along with the NOS terminator. These fragments were ligated into the BamH I and the EcoR I sites of the pBI-121 plasmid, which was prepared for cloning by cutting it with the same enzymes to release the wild type GUS cDNA. The resulting constructs (pBI-IF-GUS and pBI-OF-GUS) were subsequently digested with Hind III and EcoR I to release the DNA fragments encompassing the CaMV promoter, the IF or OF GUS cDNA, and the NOS terminator.

Finally, these fragments were ligated into the correspondent restriction sites present in the pGPTV-HPT binary vector (ATCC) to obtain the pCMV-IF-GUS-HPT and pCMV-OF-GUS-HPT binary vectors.

The resulting vectors, CMV-OF-GUS-HPT and CMV-IF-GUS-HPT now contain the CaMV35S promoter from the Cauliflower Mosaic 35 S Virus driving the GUS gene followed by a NOS terminator and polyadenylation signal (Fig. 7). In addition, this vector also contains a hygromycin resistance gene as a selectable marker that is used to select for plants containing this reporter.

#### **Generation of GUS reporter-expressing *Arabidopsis thaliana* transgenic plants.**

*Agrobacterium tumefaciens* bacteria are used to shuttle binary expression vectors into plants. To generate  $\beta$ -glucuronidase-expressing *Arabidopsis thaliana* (*A. thaliana*) plants, *Agrobacterium tumefaciens* cells (strain GV3101) were electroporated with the CMV-OF-GUS-HPT or the CMV-IF-GUS-HPT binary vector using methods known by those skilled in the art. Briefly, one-month old *A. thaliana* (ecotype Columbia) plants were infected by immersion in a solution containing 5% sucrose, 0.05% silwet and binary vector-transformed *Agrobacteria* cells for 10 seconds. These plants were then grown at 25°C under a 16 hour day and 8 hour dark photoperiod. After 4 weeks, seeds (referred to as T1) were harvested and dried for 5 days. Thirty thousands seeds from ten CMV-OF-GUS-HPT or CMV-IF-GUS-HPT-transformed plants were sown in solid Murashige and Skoog (MS) media plates in the presence of 20  $\mu$ g/ml of hygromycin (HYG). Three hundred plants were found to be HYG resistant and represented GUS expressing plants. These plants along with 300 control plants were grown in MS media for two weeks and then transferred to soil. Plants were grown for an additional four weeks under standard conditions at which time T2 seeds were harvested.

To confirm the integration and stability of the GUS vector in the plant genome, gene segregation and PCR analyses were conducted. Commonly, three out of four T1 plants transformed by *Agrobacteria* technology are expected to carry the vector inserted within a single locus and are therefore considered heterozygous for the integrated gene. Approximately 75% of the seeds (T2) generated from most of the T1 plants were found HYG-resistant and this in accordance with the expected 1:2:1 ratio of null (no GUS



containing plants), heterozygous, and homozygous plants, respectively, in self-pollinating conditions. To confirm that these plants contained the GUS expression vector, genomic DNA was isolated from leaves of T1 plants using the DNAzol-mediated technique as described above. One ng of genomic DNA was analyzed by polymerase chain reaction (PCR) to confirm the presence of the GUS vector. PCR was carried out for 25 cycles with the following parameters: 95°C for 30 seconds; 54°C for 1 minute; and 72°C for 2 minutes using primers listed above. Positive reactions were observed in DNA from CMV-OF-GUS-HPT and CMV-IF-GUS-HPT-transformed plants and not from control (uninfected) plants.

In order to assess the expression of the GUS in T1 plants, leaf tissue was collected from T1 plants, homogenized in liquid nitrogen using glass pestles, and suspended in RLT lysing buffer (Qiagen, RNeasy plant RNA extraction kit). Five micrograms of total RNA was purified according to the manufacturer's suggested protocol and then loaded onto a 1.2% agarose gel (1x MOPS buffer, 3% formaldehyde), size-fractionated by electrophoresis, and transferred onto N-Hybrid+ membrane (Amersham). Each membrane was incubated at 55°C in 10 ml of hybridization solution (North2South labeling kit, Pierce) containing 100 ng of GUS, tubulin, or HYG probes, which were generated by PCR amplification, according to the manufacturer's directions. Membranes were washed three times in 2x SSC, 0.1% SDS at 55°C, and three times in 2x SSC at ambient temperature. Detection was carried out using enhanced chemiluminescence (ECL). GUS message was detected in three out of ten analyzed transgenic lines, while no signal was found in the control plants. Collectively these studies demonstrated the generation of GUS expressing transgenic *A. thaliana* plants.

To determine the status of MMR activity in host plants, one can measure for the production of functional  $\beta$ -glucuronidase by staining plant leaves or roots *in situ* for  $\beta$ -glu activity. Briefly, plant tissue is washed twice with water and fixed in 4 mls of 0.02% glutaraldehyde for 15 minutes. Next, tissue is rinsed with water and incubated in X-glu solution [0.1M NaPO<sub>4</sub>, 2.5 mM K<sub>3</sub>Fe(CN)<sub>6</sub>, 2.5mM K<sub>4</sub>Fe(CN)<sub>6</sub>, 1.5 mM MgCl<sub>2</sub>, and 1 mg/ml X-GLU (5 bromo-4-chloro-3-indoyl-  $\beta$ -D-glucuronide sodium salt) (Gold Biotechnology)] for 6 hours at 37°C. Tissues are then washed twice in phosphate buffered saline (PBS) solution, once in 70% ethanol and incubated for 4 hours in methanol:acetone (3:1) for 8 hours

to remove chlorophyll. Tissues are then washed twice in PBS and stored in PBS with 50% glycerol. Plant tissue with functional GUS activity will stain blue.

The presence of GUS activity in CMV-IF-GUS-HPT plants indicates that the in-frame N-terminus insertion of the poly A repeat does not disrupt the GUS protein function. The CMV-OF-GUS-HPT plants treated with DMA, EMS or untreated are tested to determine if these plants produce GUS activity. The presence of GUS activity in DMA treated plants indicates that the polyA repeat was altered, therefore, resulting in a frame-restoring mutation. Agents such as EMS, which are known to damage DNA by alkylation cannot affect the stability of a polynucleotide repeat. This data indicates that plants are defective for MMR, the only process known to be responsible for MI.

These data demonstrate the utility and power of using a chemical inhibitor of MMR to generate a high degree of genetic alteration that is not capable by means of standard DNA damaging drugs. Moreover, this application teaches of the use of reporter genes such as GUS-OF in plants to monitor for the MMR activity of a plant host.

**EXAMPLE 7: Use of chemical MMR inhibitors yields microsatellite instability in microbes.**

To demonstrate the ability of chemical inhibitors to block MMR in a wide range of hosts, we employed the use of *Pichia* yeast containing a pGUS-OF reporter system similar to that described in Example 5. Briefly, the GUS-OF and GUS-IF gene, which contains a polyA repeat at the N-terminus of the protein was subcloned from the pCR-IF-GUS and pCR-OF-GUS plasmids into the EcoRI site of the pGP vector, which is a constitutively expressed yeast vector containing a zeocin resistance gene as selectable marker. pGP-GUS-IF and pGP-GUS-OF vectors were electroporated into competent *Pichia* cells using standard methods known by those skilled in the art. Cells were plated on YPD agar (10g/L yeast extract; 20 g/L peptone; 2% glucose; 1.5% bactoagar) plates containing 100 µg/ml zeocin. Recombinant yeast are then analyzed for GUS expression/function by replica plating on YPD agar plates containing 100 µg/ml zeocin plus 1 mg/ml X-glu (5-bromo-4-chloro-3-indolyl-beta-D-glucuronide sodium salt) and grown at 30°C for 16 hours. On hundred percent of yeast expressing GUS-IF were found to turn blue in the presence of the X-glu substrate while none of the control yeast

turned blue. None of the yeast containing the GUS-OF turned blue in the presence of the X-glu substrate under normal growth conditions.

To demonstrate the ability of chemicals to block MMR in yeast, GUS-OF and control cells were incubated with 300  $\mu$ M DMA, EMS, or no chemical for 48 hours. After incubation, yeast were plated on YPD-ZEO-X-GLU plates and grown at 30°C for 16 hours. After incubation, a subset of yeast expressing GUS-OF contain blue subclones, while none are seen in EMS or control cells. These data demonstrate the ability of chemicals to block MMR of microbes *in vivo* to produce subclones with new output traits.

#### EXAMPLE 8: Classes of other chemicals capable of blocking MMR *in vivo*

The discovery of anthracene compounds presents a new method for blocking MMR activity of host organisms *in vivo*. While 9,10-dimethylanthracene (DMA) was found to block MMR in cell hosts, other analogs with a similar chemical composition from this class are also claimed in this invention. These include anthracene and related analogs such as 9,10-diphenylanthracene and 9,10-di-M-tolylanthracene. Myers *et al.* ((1988) *Biochem. Biophys. Res. Commun.* 151:1441-1445) disclosed that at high concentrations, DMA acts as a potent weak mutagen, while metabolized forms of DMA are the "active" ingredients in promoting mutation. This finding suggests that metabolites of anthracene-based compounds may also act as active inhibitors of MMR *in vivo*. For instance, metabolism of anthracene and 9,10-dimethylanthracene by *Micrococcus sp.*, *Pseudomonas sp.* and *Bacillus macerans* microbes have found a number of anthracene and 9,10-dimethylanthracene metabolites are formed. These include anthracene and 9,10-dimethylanthracene cis-dihydrodiols, hydroxy-methyl-derivatives and various phenolic compounds. Bacteria metabolize hydrocarbons using the dioxygenase enzyme system, which differs from the mammalian cytochrome P-450 monooxygenase. These findings suggest the use of bacteria for biotransforming anthracene and DMA for additional MMR blocking compounds (Traczewska, T.M. *et al.* (1991) *Acta Microbiol. Pol.* 40:235-241). Metabolism studies of DMA by rat-liver microsomal preparations has found that this molecule is converted to 9-Hydroxymethyl-10-methylanthracene (9-OHMeMA) and 9,10-dihydroxymethyl-anthracene (9,10-DiOHMeA) (Lamparczyk, H.S. *et al.* (1984) *Carcinogenesis* 5:1405-1410). In addition, the trans-1,2-

dihydro-1,2-dihydroxy derivative of DMA (DMA 1,2-diol) was found to be a major metabolite as determined by chromatographic, ultraviolet (UV), nuclear magnetic resonance (NMR), and mass spectral properties. DMA 1,2-diol was also created through the oxidation of DMA in an ascorbic acid-ferrous sulfate-EDTA system. Other dihydrodiols that are formed from DMA by metabolism are the trans-1,2- and 3,4-dihydrodiols of 9-OHMeMA (9-OHMeMA 1,2-diol and 9-OHMeMA 3,4-diol) while the further metabolism of DMA 1,2-diol can yield both of these dihydrodiols. Finally, when 9-OHMeMA is further metabolized, two main metabolites are formed; one was identified as 9,10-DiOHMeA and the other appeared to be 9-OHMeMA 3,4-diol.

The metabolism of 9-methylanthracene (9-MA), 9-hydroxymethylanthracene (9-OHMA), and 9,10-dimethylanthracene (9,10-DMA) by fungus also has been reported (Cerniglia, C.E. *et al.* (1990) *Appl. Environ. Microbiol.* 56:661-668). These compounds are also useful for generating DMA derivatives capable of blocking MMR. Compounds 9-MA and 9,10-DMA are metabolized by two pathways, one involving initial hydroxylation of the methyl group(s) and the other involving epoxidation of the 1,2- and 3,4- aromatic double bond positions, followed by enzymatic hydration to form hydroxymethyl trans-dihydrodiols. For 9-MA metabolism, the major metabolites identified are trans-1,2-dihydro-1,2-dihydroxy and trans-3,4-dihydro-3,4-dihydroxy derivatives of 9-MA and 9-OHMA, whereby 9-OHMA can be further metabolized to trans-1,2- and 3,4-dihydrodiol derivatives. Circular dichroism spectral analysis revealed that the major enantiomer for each dihydrodiol was predominantly in the S,S configuration, in contrast to the predominantly R,R configuration of the trans-dihydrodiol formed by mammalian enzyme systems. These results indicate that *Caenorhabditis elegans* metabolizes methylated anthracenes in a highly stereoselective manner that is different from that reported for rat liver microsomes.

The analogs as listed above provide an example but are not limited to anthracene-derived compounds capable of eliciting MMR blockade. Additional analogs that are of potential use for blocking MMR are shown in Fig.8.

**Other classes of small molecular weight compounds that are capable of blocking MMR *in vivo*.**

MMR is a multi-step process that involves the formation of protein complexes that detect mismatched bases or altered repetitive sequences and interface these mutations with enzymes that degrade the mutant base and repair the DNA with correct nucleotides. First, mismatched DNA is recognized by the mutS heterodimeric complex consisting of MSH2 and GTBP proteins. The DNA bound mutS complex is then recognized by the mutL heterodimeric complex that consists of PMS2 and MLH1 proteins. The mutL complex is thought to interface exonucleases with the mismatched DNA site, thus initiating this specialized DNA repair process. After the mismatched bases are removed, the DNA is repaired with a polymerase.

There are several steps in the normal process that can be targeted by small molecular weight compounds to block MMR. This application teaches of these steps and the types of compounds that may be used to block this process.

#### **ATPase inhibitors:**

The finding that nonhydrolyzable forms of ATP are able to suppress MMR *in vitro* also suggest that the use for this type of compound can lead to blockade of MMR *in vivo* and mutation a host organism's genome (Galio, L. *et al.* (1999) *Nucl. Acids Res.* 27:2325-2331; Allen, D.J. *et al.* (1997) *EMBO J.* 16:4467-4476; Bjornson, K.P. *et al.* (2000) *Biochem.* 39:3176-3183). One can use a variety of screening methods described within this application to identify ATP analogs that block the ATP-dependent steps of mismatch repair *in vivo*.

#### **Nuclease inhibitors:**

The removal of mismatched bases is a required step for effective MMR (Harfe, B.D. and S. Jinks-Robertson (2000) *Ann. Rev. Genet.* 34:359-399). This suggests that compounds capable of blocking this step can lead to blockade of MMR *in vivo* and mutation a host organism's genome. One can use a variety of screening methods described within this application to identify nuclease inhibitors analogs that block the nuclease steps of mismatch repair *in vivo*. An example of the types of nuclease inhibitors are but not limited to analogs of N-Ethylmaleimide, an endonuclease inhibitor (Huang, Y.C., *et.al.* (1995) *Arch. Biochem. Biophys.* 316:485), heterodimeric adenine-chain-acridine compounds, exonulcease III

inhibitors (Belmont P, et.al., *Bioorg Med Chem Lett* (2000) 10:293-295), as well as antibiotic compounds such as Heliquinomycin, which have helicase inhibitory activity (Chino, M, et.al. *J. Antibiot. (Tokyo)* (1998) 51:480-486).

#### **Polymerase inhibitors:**

Short and long patch repair is a required step for effective MMR (Modrich, P. (1994) *Science* 266:1959-1960). This suggests that compounds capable of blocking MMR-associated polymerization can lead to blockade of MMR *in vivo* and mutation a host organism's genome. One can use a variety of screening methods described within this application to identify polymerase inhibitors analogs that block the polymerization steps of mismatch repair *in vivo*. An example of DNA polymerase inhibitors that are useful in blocking MMR activity include, but are not limited to, analogs of actinomycin D (Martin, S.J., et.al. (1990) *J. Immunol.* 145:1859), Aphidicolin (Kuwakado, K. et.al. (1993) *Biochem. Pharmacol.* 46:1909) 1-(2'-Deoxy-2'-fluoro-beta-L-arabinofuranosyl)-5-methyluracil (L-FMAU) (Kukhanova M, et.al., *Biochem Pharmacol* (1998) 55:1181-1187), and 2',3'-dideoxyribonucleoside 5'-triphosphates (ddNTPs) (Ono, K., et.al., *Biomed Pharmacother* (1984) 38:382-389).

#### **Chemical Inhibitors of Mismatch Repair Gene Expression**

MMR is a multi-protein process that requires the cooperation of several proteins such as but not limited to mutS homologs, MSH2, MSH3, MSH6, GTBP; mutL homologs PMS1, PMS2, MLH1; and exonucleases and helicases such as MutH and MutY (Harfe, B.D. and S. Jinks-Robertson (2000) *Ann. Rev. Genet.* 34:359-399; Modrich, P. (1994) *Science* 266:1959-1960). Chemicals capable of blocking the expression of these genes can lead to the blockade of MMR. An example of a chemical that is capable of blocking MMR gene expression is an oligodeoxynucleotide that can specifically bind and degrade an MMR gene message and protein production as described by Chauhan DP, et.al. (*Clin Cancer Res* (2000) 6:3827-3831). One can use a variety of screening methods described within this application to identify inhibitors that block the expression and/or function of MMR genes *in vivo*.

## DISCUSSION

The results described herein demonstrate the use of chemicals that can block mismatch repair of host organisms *in vivo* to produce genetic mutations. The results also demonstrate the use of reporter systems in host cells and organisms that are useful for screening chemicals capable of blocking MMR of the host organism. Moreover, the results demonstrate the use of chemical inhibitors to block MMR in mammalian cells, microbes, and plants to produce organisms with new output traits. The data presented herein provide novel approaches for producing genetically altered plants, microbes, and mammalian cells with output traits for commercial applications by inhibiting MMR with chemicals. This approach gives advantages over others that require the use of recombinant techniques to block MMR or to produce new output traits by expression of a foreign gene. This method will be useful in producing genetically altered host organisms for agricultural, chemical manufacturing, pharmaceutical, and environmental applications.

## PMS2 (mouse) (SEQ ID NO:14)

```

MEQTEGVST CAKAIKPIDG KSVHQICSGQ VILSLSTAVK ELIENSVDAG ATTIDRLKLD 60
YGVDLIEVSD NGCGVEEENF EGLALKHHTS KIOEFADLTQ VETFGFRGEA LSSLCALSDV 120
TISTCHGSAS VGTRLVFDHN GKITQKTPYP RPKGTTVSQV HLFYTLVPVRY KEFQRNLIKKE 180
YSKMVQVLOA YCIISAGVRV SCTNQLGQK RHAVVCTSGT SGMKENIGSV FGQQLQSLI 240
PFVQLPPSDA VCEEYGLSTS GRHKTFSTFR ASFHSARTAP GGVQQTGSFS SSIRGPVTOQ 300
RSLSLSMRFY HMVNRHQYPF VVLNVSDVSE CVDINVTPOK RQILLQEEKL LLAVLKTSLI 360
GMFDSLANKL NVNQOPLLDV EGNLVKLHTA ELEKPVVPKQ DNSPFLKSTA DEKRVASISR 420
LREAFSLHPT KEIKSRGPET AELTRSPFSE KRGVLSSYPS DVISYRGLRG SQDKLVSPFD 480
SPGDCMDREK TEKDSGLSST SAGSEEEFST PEVASSFTEE YNVSSLEDRP SQETINCGLD 540
DCRCPGTGGY LKPEDHGYQC KALPLARLSP TNAKRFKTEE RPSNVNISOR LPGQSTSA 600
EVDVAIKMNK RIVLEFSL SLAKRMKQLQ HLKAQNKHLE SYRKFRAKTC PGENQAAE 660
LRKEISKSMF AEMEILGQFN LGFIVTKLKE DLFLVDQHA DEKNFEMIQ QHTVLAQRL 720
ITPQTNLTA VNEAVLIENL EIFRKNGDFD VIDEAPVTE RAKLISLPTS KNWTFGPQDT 780
DELIFMLSDS PGVMCRFSRV RQMFASRACR KSMVIGTALN ASEMKKLITH MGEMDHPWNC 840
PHGRPTMRHV ANLDVISQN

```

## PMS2 (mouse cDNA) (SEQ ID NO:15)

```

gaattccggt gaaggtcctg aagaatttcc agattcctga gtatcattgg agggagacaga 60
taacctgtgc cagagtgatga atggtgtata tgcaacagaa atgggtgttc ctggagacgc 120
gtcttctccc facagcgcca ccgcaactct cccgcggtga ctgtgactgag aggaactcctg 180
catccatgag gccaaaccgaa ggcgtgagta cagaatgtgc taaggccatc aagcctattg 240
atgggaagtc agtcacgtca attgttctg ggcagggtgat actcagttta agcacagtg 300
tgaaggagtt gatagaaaa agttagatgt ctggtgtcac tactattgat ctaaggctta 360
aagactatgg ggtggacctc attgaagttt cagacaatgg atgtggggtg gaagaaagaa 420
actttgaagg tctagctctg aaacatcaca catctaagat tcaagagttt gccgacctca 480
cgccpgtggc aactttcggc tttcgggggg aagctctgag ctctctgtgt gcactaagtg 540
atgtcactat atctacctgc caacgggtctg caacggttgg gactcgaact gtgtttgacc 600
ataatgggaa aatcacccag aaaactccct acccccgacc taaaggaacc acagtcaagt 660
tgcagcactt attttatata ctaccctgct gttacaaaga gtttcagagg aacattaaaa 720

```

aggagatctt ccaaatgggt caggctcttc aggcgtactg tatcatctca gcaggcgctc 780  
 gtgttaagctg cactaatcag ctgggacagg ggaagcgcca cgctgtgtgtg tgcacaaagg 840  
 gcacgtctcg catgaaggaa aatatcggtg ctgtgtgttg ccagaagcag ttgcaaaaggc 900  
 tcattctctt tgttcagctg ccccttagtg acgctgtgtg tgaagagtac ggctcgagca 960  
 ctccaggacg ccacaaaacc tttctacgt ttcgggcttc attccacagt ccgcgcagcg 1020  
 cgccgggagg agtgcaacag acaggcagtt tttctctate aatcagagcg cctgtgacct 1080  
 agcaaaagctg tctaaagctg tcaatgaggt tttatcacat gtataaccgg catcagtaacc 1140  
 catttgctgt ccttaacgtt tccgttgact cagaatgtgt ggaatattaat gtaactccag 1200  
 ataaaaggca aattctacta caagaagaga agctattgct ggccgtttaa aagacctcct 1260  
 tgataggaat gtttgacagt gatgcaaaaca agcttaagt caaccagcag ccactgctag 1320  
 atgttgaagg taacttagta aagctgcata ctgcagaact agaaaagcct gtgccaggaa 1380  
 agcaagataa ctctccttca ctgaagagca cagcagacga gaaaagggtta gcatccatct 1440  
 ccaggctgag agaggccttt tctctctate ctactaaaaga gatcaagctc aggggtccag 1500  
 agactgctga actgacacgg agttttccaa gtgagaaaag gggcgctgtt tctctctate 1560  
 ctccagactg catctctaac agaggcctcc gtggctcgca ggacaaattg gtgagtcaca 1620  
 cggacagccc tgtgtgactg atggacagag agaaaaataga aaaagactca gggctcagca 1680  
 gcacctcagc tggctctcag gaagagtcca gcccccagca agtgggccagt agctttagca 1740  
 gtgactataa cgtgagctcg ctagaagaca gaccttctca gaaaagccta aactgtgtgtg 1800  
 acctggaact ccgtctctca ggtacagga agtctctgaa gcccaagacg catggatatt 1860  
 aatgcaaaag tctactctta gctcgtctgt caccaccaaa tgccaagcgc tcaagagacag 1920  
 aggaaaagcc ctcaaatgtt aacatttctc aaagattgct tggctctcag agcacctcag 1980  
 cagctgagct cgtatgtacc ataaaaatga ataagagata cgtgctctc gagtctctc 2040  
 tgagtctctc agctaagcga atgaagcagt tacagcacct aaaggccgag acaaaacatc 2100  
 aactgagttc cagaaaaatt agggccaaga tttcgctcgg gaaaaccacaa gaccgagaag 2160  
 atgaactcag aaaaagagtt agtaaatcga tgtttgacga gatggagatc ttgggtcagt 2220  
 ttaactggg atttatagta accaaaactga aagaggacct ctctcgtgtg gaccagcagt 2280  
 ctgcgagatg gaagtacaac tttgagatgc tgcagcagca caccggtgctc caggcgatga 2340  
 ggctcatcac accccagact ctgaacttaa ctgctctcaa tgaagtctga ctatagaaa 2400  
 atctggaatt atctcagaag aatggccttg actttgtcat tgatgagagt gctccagta 2460  
 ctgaaaggcg taaattgatt tcttaccaca ctatgaaaaa ctggaccttt ggaccocaaag 2520  
 atataagatg actgatcttt atgttaagt acagccctgg ggtcatgtgc cggccctcac 2580  
 gagtacagca gatgtttgct tccagagcct gcggaagtc agtgatgatt ggaacggcgc 2640  
 tcaatgcgag cgagatgaag aagctcatca ccccatcggg tgagatggaac caccctcaga 2700  
 actgccccca cggcaggcca accatgaggc acgttgccaa tctctcagat ctctcaga 2760  
 actgacacac ccttctagc atagagttaa ttacagattg ttccggtttgc aaagagaagg 2820  
 ttttaagtta tctgattatc gttgtacaaa aattagcatg ctgcttttaa ctctgagta 2880  
 catttaaaag cagtggttaa gcaggcatga ttgagtgctc ctctagctca gctactgtgg 2940  
 tgatcgggtg ggagctgtg tgagccagc actttgagac cactccagcg cacattcatg 3000  
 agactcaatt caaggacaaa aaaaaaaaga tatttttgaa gccttttaaa aaaaaa 3056

## PMS2 (human) (SEQ ID NO:16)

MERAESSSTE PAKAIKPIDR KSVHIQCSGQ VVLSLSTAVK ELVENSIDAG ATNIDLKLDK 60  
 YGVDLILEVD NGCGVEEENF EGLTLKHHTS KIQEFADLTG VETFGFGEA LSSLCALSDV 120  
 TISTCHASAK VGTRLMFHDH GKIIQTPYP PRGTTVSVO QLESTLVRH KEFQRIKKE 180  
 YAKMVQVLHA YCIISAGIRV SCNQLGQCKG RQPVCTCGVS PSIKEMIGSV FGQKQLQSLI 240  
 PFYQLPVSLS VCEEYGLSCS DALHNLFYIS GFISQCTHCV GRSDTORPFF FINRCPDFA 300  
 KVCRLVENVY HMYNRHQYFF VVLNISVDSE CVDINVTPDK ROILLQEELK LLAVLTSTRI 360  
 GMFDSOVNKL NVSQPIILDV EGNLIRKMHAA DELEPMVEKQ DQSPSLRTGE EKKDVSISRL 420  
 REAFSLRHHT ENKPHSPKPT EPRRSPLGOK RGMLSSTSTG AISPDKVLRF QKVASSSHG 480  
 PSDPTORAEV EKDSGHGSTS VDESGFSIPD TGSCHSSEYA SSPCDGRSGO EHDVSQKAP 540  
 ETDDSFSDVD CHSNQEDTGC KFRVLPQPTN LATPNTKREK KEILSSSDI COKLVNTQDM 600  
 SASQVDAVAK INKKVQLLDV SMSSLAKRIK QLHHEAQOSE EQONYRKFFA KICPGENQAA 660  
 EDELRLKISK TMFAEMEIIQ QFNLFGIITK LNEIDFIVDQ HATERKYNEF MLQQHTVLCG 720  
 QRLIAQTINL LTVANEAIV ENLEIFRKNQ DFVIDENAP VTDEKILSL PTKNWTFGP 780  
 QDVELIELML SDSPGVMCPR SRVKQMFASR ACKRSVMIGT ALNTSEMKKL ITHMGEMDHF 840  
 WNCFHGRPTM RHIANLGVIS QN 862

## PMS2 (human cDNA) (SEQ ID NO:17)

cgaggcgat cggtgtgtg atccatggag cgagctgaga gctcgagtag agaactgct 60  
 aaggccatca aacctattga tcggaagtca gtccatcaga ttgctctgg gcaggtggta 120



ctgagtcctaa gcactgcggt aaaggagtt gtagaaaaca gtcgtgatgc tgggtgccact 180  
 aatattgatc taaagcttaa ggactatgga gtggatctta ttgaagtttc agacaatgga 240  
 tgtggggatg cgaagctaaa cttcgaagcg ttaactctga aacatcacac atctaaagatt 300  
 caagagtttg ccgaactaac tccaggttga acttttggct ttcgggggga agctcttgagc 360  
 tcaactttgt cactgtagca cactgaccatt tctacctgcc acgcatcggc gaaggttggga 420  
 actgactgca tgtttgatca caatggggaaa attatccaga aaacccccta ccccccccc 480  
 agaggagacca cagtcacgct gcagcagttg attatccaga tacccttgcc ccataaggaa 540  
 ttccaaagag atataaagaa ggagtatgcc aaaaatggcc aggtcttaca tgcactagct 600  
 atcatttcaa caggcatccg tgaagtgtgc accaatcagc ttggacaagg tgtgtttggg 720  
 cctgtgtgat gcacaggtgg cacttctttt gttcagctgc cccctagtga ctcctgtgtg 780  
 cagaagcagt gttttagctt ttcggatgct ctgcataatc tttttacat ctcaggtttc 840  
 gaagagtacg gcacgcgatg agttggaagg agttcaacag acagacagtt tttctttatc 900  
 atttcacaa aaccggcgcg agcaaaaggt ctcagactcg tgaatgaggt caccacatg 960  
 tataatcgag gataatcatt ttaactccga ttatctccga gcttttgggt 1020  
 ctactcctaa ttactccttt gataggaatt tttgtatgt aagaggaaaa gcttttgggt 1080  
 agtcagcagc cactgcttga tgggtgaagg tcccttctac attttgctac atgtcaacaa gctaaatgtg 1140  
 gaaaaagcca cactgcttga tgggtgaagg tcccttctac attttgctac atgtcaacaa gctaaatgtg 1200  
 aaagcagctg ccatttctca actgcgagag cacttcttct tctgcacac taaggagctg aagagaaaa 1260  
 aagccttcca gcccacaagc tccagaacca agaaggagcg tctgcacac taaggagctg aagagaaaa 1320  
 atgctgtctt gtcgactctc aggtgccatc cctgcacaa gctctagtag acctcagaaa 1380  
 gaggcagctt gttccagctc cggaccaggt gaccctacgt acagacgcca ggtggagaaa 1440  
 gactcggggc acggcagcac ttcogtggat tctgaggggt tcagcatccc agacacggcg 1500  
 agtcactgca gcagcgagta tgcggccagg tccccagggg acagggggctc gcaggaacat 1620  
 gtggactctc agagaaaaag ccctgaaact gacgactct tttcagatgt ggaactgcat 1680  
 tcaaaacagg aagataccgg atgtaaaatt cgaattttgc cctcagccaa cttctctcga 1740  
 accccaacca caaagcgttt taaaaaagaa gaaattcttt ccatgtctga ctttcttcaa 1800  
 aagttagtata atactcagga catgtcagcc tctcaggttg tctcaggttg tctcaggttg 1860  
 aagaaaagtt tgcccttgat ctttcttatg agttctttag ctaaaagcat aaagcagtaa 1920  
 catcatggaag cacagcaaa ggaaggggaa cagaattaca ggaagttagt gccaagaagt 1980  
 tgtcctggag aaaaatcagc agccgaagat gaactaaaga aagagataag taaaaacgat 2040  
 ttgacagaaa tggaaatcat tggtcagttt aaactcggat ttataataac caaatcgaat 2100  
 gaggatctat tcalagtggg ccagcatgcc acggacgaga agtataact cgaattctga 2160  
 cagcagcaca ccgtgtctca ggggcagagg ctcatagcac ctcagactct ccaacttaact 2220  
 gctgttaatg aagctgtct gatagaaaat ctggaaatat ttgaaaagaa ttagaaagaa 2280  
 ttgttatcgt atgaaaatgc tccagtcact gaaagggcta aactgatctc cttgcacact 2340  
 agtaaaactc ggactctcgg accccaggag gtcgatgaa cgtcatgact gctgagcgac 2400  
 agccctgggg tcatgtgcgc gccctccccg gtaacagcag gctgttgcctc cagagcctgc 2460  
 cggagtcggg tgaatgattg gactgtctct aacacaagg agatgagaaa catgatcacc 2520  
 caatggggg agatggagca cccctggaa tgtccccatg gaagggccaa catgagacac 2580  
 atcgccacc ttgggtgtcat ttctcagaa tgaccgtagt cactgtatgt aat aat tgg 2640  
 ttatogcag atttttatgt ttgaaagac agagtcttca ctaaccttt tttgttttaa 2700  
 atgaaactg ctactttaaa aaaaatacaca tcaaccccat ttaaaagatg tttgtgagaac 2760  
 cttttcaaac c 2771

## PMS1 (human) (SEQ ID NO:18)

MKQLPAATVR LLSSSSQIITS VSVVKELIE NSLDAGATSV DVKLENYGFD KIEVRNNGEG 60  
 IKAVDAPVMA MKYYTSSKINS HEDLENLTYY GFRGEGALSI CCAIEVLIT RTADNFSTQ 120  
 VYLDGSPGHIL SQKPSHLGGQ TTVTALRLFK NLVPRKQFYS TARKCKDEIK KIQDLMSFG 180  
 ILKPDRIIVF VHNKAVIWKQ SRVSDHKMAL MSVLGTAVMN NMESFYHSE ESQYILSGFL 240  
 PKCDADHSFT SLSTPERSFI FINSRPVHOK DILKLRHHY NLKCLKESTR LVYVFFLKID 300  
 VPTADVDVNL TPDKSVQLLQ NKESVLIAL E NLMTTCYGLP PSTNSYENNK TDVSAADIVL 360  
 SKTAETDVNL NKVESSEKGY SNVDTSVIF QNDMHNDESG KNTDDCLNQG ISIGDFGKYN 420  
 CSSEISINDK NTKNAQDLS MSNVSWNSQ TEYSKTCFIS SVKHTQSENG NKPHIDESGE 480  
 NEEEAENLDS SEISADEWSR GNILKNSVGE NIEPVKILPV EKSLPCKVNS NNPYIEPQNM 540  
 LNEDSCKNKS NVINDKSGKV TAYDILLSNRV IKKPMASAL FVDQHRPQFL IENPKTSLED 600  
 ATLQIEELWK TLSEEEKLYK EKAATKDLER YNSQMKRAIE QESQMSLKDG RKKIKPTSAW 660  
 NLAQKHKLKT SLNQPWLME LLQSQIEKRR SQNIKMWQIP FSKMKNLKFN KKNQNMVLEE 720  
 KDEPCLIHNL RFPDADWLMTS KTEVMLLNYP RVEEALLFKR LLENHKLPAE PLEKPMITLE 780  
 SLFNGSHYLD VLYKMTADDO RYSGSTYLSL PRLTANGFKI KLIPGVSITE NYLIEGEMAN 840

CLPFYGVADL KEILNAILNR NAVEVYECRP RKVISYLEGE AVRLSRQLPM YLSKEDIQDI 900  
IYRMKHQFGN EIKECVHGRF FFHHLTYLPE TT 932

## PMS1 (human) (SEQ ID NO:19)

ggcacgagtg gctgcttgctg gctagtggat ggtaattgcc tgcctcgccg tagcagcaag 60  
ctgctctgtt aaaaagcgaaa atgaacaacat tgcctgcggc aacagtttga ctcctttcaa 120  
gttctcagat catcactctcg gtggtcagtg ttgtaaaaga gcttatttga aactcctgtg 180  
atgctggtgc cacaagcgta gatgttaaac tgggagaacta tggatttgat aaaaattgag 240  
tgcgagataa cggggaggggt atcaaggctg ttgatgcacc tgtaatggca atgaagtact 300  
acacctcaaa aataaatagt catgaagatc ttgaaaaattt gacaacttcc ggttttcgtg 360  
gagaagcctt ggggtcaatt tgggtgtatg ctgaggtttt aattacaaca agaagcgctg 420  
ctgataattt tagcaaccag tatgttttag atggcagtg ccacatactt tctcagaaac 480  
cttcacatct tgggtcaagg acaactgtaa ctgctttaag attatttaag aatctacctg 540  
taagaaaagca gttttactca actgcacaaa aatgtaaaag tgaataaaa aagatccaag 600  
atctcctcat gagctttgtg atccttaaac ctgacttaag gatgtctctt gtacataaca 660  
aggcagttat ttggcagaaa agcagagtat cagatcacaa gatggctctc atgtcagttc 720  
tggggactgc ttgtatggaat aatatgggat cctttcagta ccactctgaa gaactctaga 780  
tttatctcag tggatttctt ccaaaagtgt atgcagacca ctctttcaat agtctttcaa 840  
ctcagaaaag aagtttcatc ttcalaacaa ctgcacactc acatcaaaaa gatactttaa 900  
agtttaattc gaaactatca aatctgaaat gcctaaaagg atctactcgt atctactcgt 960  
tttcttctc aaaaactcgat gtctctacag ctgatgttga tgtaaattta acaccagata 1020  
aaagccaagt attattacaa aataaggaal ctgttttaag tgctcttgaa aactgatga 1080  
cgactgttta cttagtaaca atctcttaca aaataataaaa acagatgttt 1140  
cgcgagctga atcagatttt agtaaaacag cagaacacaga tgtgtctttt aataaagatg 1200  
aatcatctgg aagaatttgc tcaaatgttg atactctcgt catctccatt caaaaatgta 1260  
tgcaataatg tgaactctga aaaaacactg atgatttgtt aaatcaccag ataatgtatt 1320  
gtgactttgg ttatggtcat tgtagtagtg tatcatgtta catgtgataa aacataaaga 1380  
atgactttca ggaactttca atgagtaagt tatcatggga gaactctcag acggaataa 1440  
tataaacttg ttttataagt tccgtttaacg acaccagtc agaaaatggc aataaagacc 1500  
atataagta gagtggggaa aatgaggaag aagcaggtct tgaaaactct tcggaaattt 1560  
gtggagcag gtggagcagc ggaataatc tttaaacttc agtgggagag aatattgaac 1620  
ctgtgaaat tttagtctg gaaaaaagt taccatgtaa agtaagtaat aataattatc 1680  
caatccctga acaaatgaat cttaatgaag attcatgtaa caaaaaatca aatgtaatg 1740  
ataaaaaatc tggaaaagtt acagctttat atttactag caatcgagta atcaagaaac 1800  
ccatgtcagc aagtgtcttt tttgttcaag atcatcgctc tcaagtttcc acatgtagtg 1860  
ctaagctag tttagaggaat gcaacactac aaattgaaag actgtggaag acattgagtg 1920  
aagagaagaa agctgaatat gaagagaagc ctactaaaag cttgggaacg agaaaatgca 1980  
aaatgaagag accgattgaa caggagtcac aaatgtcaat aatgtgagc gtaaaaagc 2040  
taaaaaccac cagcgcgatg aatttggccc agaaagcaca atgaataacc tcaattacta 2100  
atcaacaaa ctctctcagt cctctcagc ccaaaaattg aaaaagaagg agcaaaaaa 2160  
ttaaaattgt acagatcccc ttttctatga aaaacttaaa aataaatttc agtaacaaa 2220  
acaaaattga cttagaagag aaggatgaac ctgcttgat ccacaatctc aggtttctcg 2280  
atgcatgtgt aatgacattc aaaacagagc taatgttatt aaatccatat agagtagaag 2340  
aacccctctt atttaaaaa ctctctgaga atcaaaaaat tctcgagag cactcggaag 2400  
agccaattat gttaaocagag agtcttttta atggtactca tttattagac gtttttata 2460  
aaatgacacg agatgaccaa agatcacagt gatcaacta cctgtctgat cctgcttata 2520  
cagcgaatg tttaagaata aaattgatac caggagtttc aattactgat aactactgtg 2580  
aaatagaagg aatggctaat tgtctcccat tctatggagt agaagattt aaagaacttg 2640  
ttaatgctat attaaacaga aatgcaaaag aagtttatga atgtagacat cgaaaagta 2700  
taagtatttt agagggagaa cgagtcgctc tatccagaca attaccatg tacttattca 2760  
aagagaacat ccaagacatt atctacagaa tgaagacaca gtttggaaat gaaattaaag 2820  
agtgtgttca tggctgccca ttttttcatc attaacctta tcttcagaa actacatgat 2880  
taaatattgt taagaagatt agttaccatt ctgtcataaa acagcatgat 2940  
tctgttttca aattatcttt gtattatgtg tccactgtgt attttttaaa tcagacttca 3000  
ctgactttgt ttttatattg aaaaagttcc acgtatttga gaaaacgtaa ataaactaat 3060  
aac 3063

## MSH2 (human) (SEQ ID NO:20)

MAVQPKETIQ LESAAEVGFV RFFQGMPEKP TTTVRLFRDG DFYTHAGEDA LLAAREVFKT 60

QGVIKYMGPA	GAKNQLSVL	SKMNFESVK	DLLLVQRQYV	EVYKNRAGNK	ASKENDWYLA	120
YKASPNGLSQ	FEDILFGNND	MSASIGVVGV	KMSAVDGRQ	VGVGYVDSIQ	RKLGLCEFPD	180
NDQFSNLEAL	LQIQPKCEV	LPGGETAGDM	GKLRQIIQRG	GILITERRKA	DFSTCKDIYQD	240
LNRLRLKQK	BQMSAVLPE	MENQVAVSSL	SAVIKLELL	SDSNFGQFE	LTTFDSQYM	300
KLDIAAVRAL	NLFQGSVET	TGSQSLAALL	NKCKTPQQR	LVNQWIKQPL	MDKNRIEERL	360
NLVEAFVEDA	LRRTLQEDL	LRFPDLNRL	AKKFQQAAN	LQDCYRLYOG	INQLPNVIOA	420
LEKHGKHOK	LLLAIVFTSL	TDLRSDFSXF	QEMIEITLDM	DQVENHEFLV	KPSFDPNLSE	480
LREIMNDLEK	KMQSTLISA	RDGLDGPQG	IKLDSSAQFG	YVFRVTCKEE	KVLNRNKNFS	540
TVDIQKNGVK	FTNSKLTSLN	EYYTKNKTEY	EEAQDAIVKE	IVNISSGYVE	PMQTLNDVLA	600
QDIAVVSFAH	VSNQAPVYV	RPAILEKGQG	RIILKASRHA	CVEVQDEIAF	IPNDVYFEKD	660
KQMFHIIITGP	NMGKSTYIR	QTGVIVLMAQ	IGCFVPCESA	EVSIIVDCILA	RVGAGDSQLK	720
GVSTFMAEML	ETASILRSAT	KDSLIIIDEL	GRGTSTYDGF	GLAWAISEYI	ATKIGAFCMF	780
ATHFHELTAL	ANQIPTVNNL	HVTALTTEET	LTMLYQVKKG	VCDQSGFLHV	AELANFPKHV	840
IECAKQKALE	LEEFQYIGES	QYDIMEPA	KKCYLEREQG	EKIQEFLSK	VKQMPFTEMS	900
EENITTKLKG	LKAEVIAKN	SFVNEIISRI	KVTT			934

## MSH2 (human cDNA) (SEQ ID NO:21)

ggcgggaaac	agcttagtg	gtgtggggtc	gcgcattttc	ttcaaccagg	aggtgaggag	60
gtttcgacat	ggcggctgcag	ccgaaggaga	cgctgcagtt	ggagagcgcg	gcgcaggtgc	120
gcttcctgcg	gctctcttcag	ggcatgccgg	agaagccgac	caccacagtg	cgctcttcgc	180
accgggggca	gctctatcag	gcgcacggcg	aggacgcgt	gctggccgcc	cgggaggtgt	240
tcaagacaga	gggggtgtatc	aagtaacatg	ggcggcgagg	agcaagaagt	ctgcagaggt	300
ttgtgcttag	taaaatgaat	tttgaatctt	ttgtaaaaaa	tcttctctgt	gttcgtcagt	360
atagagttag	agctttataag	aataagacgt	gaataaaggc	atccaaaggag	aatgattggc	420
atattggcata	taaggctctc	cctggcaact	tctctcagtt	tgaagacatt	ctcttttgta	480
acaatgatat	gtcagctctc	atgtgtgtgt	tgggtgttaa	aatgtcccca	gttgatggcc	540
agagacaggt	tggagtgtgg	tatgtggatt	ccatacagag	gaacactagga	ctgtgtgaat	600
tccctgataa	tgcactgttc	tccaactctg	aggctctcct	catccagatt	ggaccacagg	660
aattgtgttt	accggggagg	gagactgctg	gagacatggg	gaacactgaga	cagataattc	720
aaagaggagg	aattctgata	acagaaaaga	aaaaagctga	ctttccacca	aaagacattt	780
atcaggacct	caaccgggtt	ttgaaaggca	aaaaggaggc	gcagatgaat	agtgcgtgat	840
tgcagaaat	ggagaaatcag	gttgcaagtt	catcactgtc	tgcggtaact	aagtttttag	900
aactcttatc	atagatttcc	aactttggac	agtttgaact	gactactttt	gactctcagc	960
agtatatgaa	ttgtgatatt	gcagcagtc	gagcccttaa	cctttttcag	cttctctgtt	1020
aagatcacac	ttgctctcag	tctctggcgt	ccttgctgaa	taagtgtaaa	accctcagg	1080
gacaaagact	ttgtaaccag	tggattaagc	agcctctcat	ggataagaaa	agaaatcagg	1140
agagattgaa	tttagtggaa	gcttttgtat	aagatgcaga	attgaggcag	actttacaag	1200
aagattttact	tcgtcgattc	ccagatctta	accgacttgc	caagaagttt	caagaagcag	1260
cagcaaaactt	acaagattgt	taccgaactc	atcagggtat	aaatcaacta	ctctaattga	1320
tacaggtctc	ggaaaaaact	gaaggaaaac	accagaaaat	ctttttggca	gttttgtgta	1380
ctctctctac	tcatctctgc	tctgacttct	ccaagtttca	ggaaaatgata	gaacaaactt	1440
tagatatgga	tcaggtggaa	aaccatgaat	tctttgtaaa	acctcttatt	aaatgcattc	1500
tcagtgaaat	aagagaaaata	atgaatgact	tggaaaagaa	gatgcagtc	acattataaa	1560
gtgcagccag	agatctttgc	tggaccctgt	gcaaacagat	taaacctggat	tcacagtcac	1620
agtttggtata	ttacttttcg	gtaaacctgt	aggaagaaaa	agtctctcgt	aacaataaaa	1680
acttttagtac	tgtatgatac	cagaagaatg	gtgttaaat	taccacagat	aaatgcattt	1740
cttttaaatg	agagataacc	aaaaataaaa	cagaataatg	agaagcccgag	gatgccattt	1800
tcaaaagaat	tgtcaaatat	tcttcaggct	atgtagaacc	aatgcagaca	ctcaatgatg	1860
gttgtagctca	gtgatgatgt	gttgtcagct	ttgtgcacgt	gtcaaatgga	gcacctgttc	1920
catatgtatc	accagccatt	ttggagaaa	gacaaggagg	aattatata	gaactcattc	1980
ggcatgctgt	tgttgaaagt	caagatgaaa	ttgcatttat	tcttaatgag	gtataacttt	2040
aaaaagataa	acagatgttc	cacatcatta	ctggcccca	tggtggaggt	aaatacaact	2100
atatttcgaca	aactgggggt	atagactcca	tggcccaaat	tggtgttttt	gtgccattgt	2160
agtcagacaga	agttgcattt	gtggactgca	tcttagcccc	agtagggggt	gtgacagtc	2220
aattgaaagg	agttctccag	tcatgtgctg	aaatgttggg	aactgctctt	atccctcaggt	2280
ctgcacacaa	agattcatta	ataactcatg	atgaattggg	aagaggaact	tctacctact	2340
atggattttg	gttagcatgt	gctatattca	aatacattgc	aacaaagatt	ggtgcttttt	2400
gcattgtttgc	aaccattttt	catgaactta	ctgcctttgc	ccactcagata	ccactcgtta	2460
ataactctaca	tgtcacagca	ctcaccactg	aaagagacct	aactatgctt	tatcaggtga	2520

agaaaggtgt	ctgtgatcaa	agttttggga	ttcatgttgc	agagcttgct	aatttcctta	2580
agcatgtaat	agagtggtgt	aaacagaaag	ccctgggaact	tgaggagttt	cagtatattg	2640
gagaatcgca	aggatatgat	atcatggaa	cagcagcga	gaagtgtcat	ctggaaagag	2700
agcaaggtga	aaaaattatt	caggagttcc	tgccaaggt	aaaacaagt	ccctttactg	2760
aaatgtcaga	agaaaacatc	acaataaagt	taaaaacagt	aaaagctgaa	gtaaatgcaa	2820
agaaataatg	ctttgtgaat	gaaatcattt	cacgaataaa	agttactacg	tgaaaatccc	2880
cagtaataag	atgaaggtaa	tattgataag	ctattgtctg	taaatgtttt	atattgtttt	2940
atattaaccc	tttttccata	gtgttaactg	tcagtgtccc	tgggtcatca	acttaataag	3000
atatttagta	atattttact	ttgaggacat	tttcaaaagt	ttttattttg	aaaatgtaga	3060
ctgtgaactg	aggagctgtt	gcaattgaca	taggcaataa	taagtgtatg	gctgaatttt	3120
ataaatataa	tcagtgtagt	tggtg				3145

## MLH1 (human) (SEQ ID NO:22)

MSFVAGVIRR	LDETVVNRIA	AGEVIQRPAN	AIKEMIENCL	DAKSTSIQVI	VKEGGLKLQ	60
IQONGTGIRK	EDLDIVCERF	TTSKLQSFED	LASISTYGRF	GEALASISHV	AHVTITTTKA	120
DGKCAVRASY	SDGKLKAPPK	PCAGNQGTQI	TVEDLFYNIA	TRRKALKNPS	EYEGKILEVV	180
GRYSVHNAGI	SFSVKKQGET	VADVRLPNA	STVDNIRISF	GNVNSRELIE	IGCEDKTLAF	240
KMNGYISNAN	YSVKKCIFLL	FINHRLVEST	SLRKAETVY	AAFLPKNTHP	FLYLSLEISF	300
QNVVDVNHPT	KHEVHFLEHE	SILERVQOHI	ESKLLGSNS	RMVFTQTLPL	GLAGPSGEMV	360
KSTTSLTSSS	TSGSSDKVYA	HQMVRTDSRE	KQLDAFLQPL	SKPLSSQPOA	IVTEKTDID	420
SGRARQDEE	MLELPAPAEV	AAKNQSLQED	TTKGTSEMS	KRGPTSSNPR	KRHREDSVE	480
MVEDDSRKEM	TAACFPERRI	INLTSLVSLQ	EEINEQGEHV	LREMLHNHSF	VGCNVQWAL	540
AQHQTLLYLL	NTTKLSEELF	YQILYDFAN	FGVLRLSEPA	PLFDLMLAL	DSPEGWTEE	600
DGPKLEGLAE	IVEFLKKKAE	MLADYFSLEI	DEEGLNLGLP	LLIDNVVPLP	EGLFIFIFPL	660
ATEVNNDEEK	ECFESLSKEC	AMFYISIRKQY	ISEESTLSGQ	QSEVPGSIPN	SKWKTVEHIV	720
YKALRSHILP	PKHFTEDGNI	LQLANLPDLY	KVFERC			756

## MLH1 (human) (SEQ ID NO:23)

cttggtctct	ctggcgccaa	aatgtcgttc	gtggcagggg	ttattcggcg	gctggacgag	60
acagtgggtg	accgcatcgc	ggcgggggaa	gttatccagc	ggccaggtaa	tgctatcaaa	120
gagatgattg	agaactgttt	agatgcaaaa	tcacaaagta	ttcaagtgat	tgttaaaagag	180
ggaggccctga	agttgattca	gatccaagac	aatggccacc	ggatcaggaa	agaagatctg	240
gatatttgat	gtgaaaagtt	cactactagt	aaactgcagt	cctttgagga	tttagccagt	300
atttctacct	atggcctttc	aggtgaggct	ttggccagca	taagccatgt	ggctcatggt	360
actattacc	cgaaaaacagc	tgatggaaag	tggtcataca	gagcaaggtta	ctcagatgga	420
aaactgaaag	ccctctctaa	accatgtgct	ggcaatcaag	ggaccagcat	cacggtgtgag	480
gacctttttt	acaaacatagc	cacgaggaga	aaagctttaa	aaaatccaag	tgaagaatat	540
gggaaaaatt	ttggaagttg	ttgcaggtat	tcagtaacaa	atgcaggcat	tagtttctca	600
gttaaaaaac	aagggaagac	agtagctgat	gttaggcac	taccacatgc	ctcaaccgtg	660
gacaaatttc	gctccatctt	ttggaatgct	gttagtcgag	aactgataga	taattgtgtg	720
gaggataaaa	ccctgacctt	caaaatgaat	ggttatcatat	ccaatgcaaa	ctactcagtg	780
agaagaatgca	cttcttacct	cttcatcaac	catcgtcttg	tgaccacca	ttccttgaga	840
aaagccatag	aaaacagtta	tgacgacctt	ttgcccacaa	acacacaccc	attcctgtac	900
ctcagtttag	ccagaatgtg	ccagaatgtg	gatgttaagt	tgaccacca	aaagcatgaa	960
gttcactgttc	tgacagaggga	gagcatcctg	gagcggtgtg	agcagcacat	cagagagcaag	1020
ctcctgtgct	ccaatctctc	caggatgtac	ttcacccaga	ctttgtctac	ggagcttgct	1080
ggccccctctg	gggagatggt	taaatccaca	acaagtctga	cctcgtcttc	tacttctgga	1140
agtagtgata	aggtctatgc	ccaccagatg	gttctgacag	attccccgga	acagaaagctt	1200
gatgtcattc	tgacgacctt	gagcaaaaccc	ctgtccagtg	agccccaggg	catgtctaca	1260
gaggaataaga	cagatatttc	tagtggcagg	gctaggcagc	agagtaggga	gaagtgtgaa	1320
ctccccagccc	ctgctggaat	ggctgccaaa	aatcagagct	tgaggggggga	tacaaacaaag	1380
gggactctcag	ctgatgtcaga	gaagagagga	cctacttcca	gcaaccccag	aaagagacat	1440
cgggaagatt	ctgagtggtga	aatggttgaa	gatgatctcc	gaagaggaat	gactgcagctt	1500
atgccccccc	ggagaaggat	cattaacctc	actagtgttc	tgaatgtcca	ggaagaaatt	1560
atgagcagc	gacatgaggt	ctcccgagg	atgttgcata	accactctct	cgtaggctgt	1620
gtgaatctc	agtggtgtc	ggcagcagat	caaaacaaat	tatacctctt	caacaccacc	1680
acgttagtg	agaagactgt	ctaccagata	ctcatttatg	attttgccaa	ttttggtgtt	1740
ctcaggttat	cggagccagc	accgctcttc	gaccttgcca	tgctgtcctt	agatagctca	1800

## MOR-0017

## PATENT

gagagtggct	ggacagagga	agatgggtccc	aaagaaggac	ttgctgaata	cattgttgag	1860
ttctgaaga	agaagcctga	gatgcttgca	gactattttc	ctttggaaat	tgatgaggaa	1920
gggaacctga	ttggtattacc	ccttctgatt	gacaactatg	tgcccccttt	ggagggagctg	1980
cctatcttca	ttcttcgact	agccactgag	gtgaattggg	acgaagaaaa	ggaatgtttt	2040
gaaagcctca	gtaaagaatg	cgctatgttc	tattccatcc	ggaagcagta	catatctgag	2100
gagtcagccc	tctcaggcca	gcagagtgaa	gtgcctggct	ccattccaaa	ctcctggaga	2160
tggactgtgag	aacacattgt	ctataaagcc	ttgcgctcac	acattctgcc	tccttaaacat	2220
ttcacagaag	atggaataat	ccctgcagctt	gctaaccctg	ctgatctata	caaaagtctt	2280
gagaggtggt	aaatatggtt	atttatgcac	tgtgggatgt	gttctctctt	ctctgtattc	2340
cgatacaaa	tgttgtatca	aagtgtgata	tacaaaagtg	accaaataaa	gtgtgtgtag	2400
cacttaagac	ttatccttgc	cttctgatag	tattccttta	tacacagtg	attgattata	2460
aataaataga	tgtgtcttaa	cata				2484

## hPMS2-134 (human) (SEQ ID NO:24)

MERAESSSTE	PAKAIPIDR	KSVHQICSGQ	VVLSLSTAVK	ELVENS LDAG	ATNIDLKLDK	60
YGVDLIEVSD	NGCGVEEENF	EGLTLKHHTS	KIQEFADLTQ	VETFGFRGEA	LSSLCALSDV	120
TISTCHASAK	VGT					133

## hPMS2-134 (human cDNA) (SEQ ID NO:25)

cgaggcgagt	cggtgtgttc	atccatggag	cgagctgaga	gctcgagtac	agaacctgct	60
aaggcatcga	aacctattga	tcggaagtca	gtccatcaga	tttgctctgg	cgaggtggta	120
ctgagctctaa	gcactcgggt	aaaggagtta	gtagaaaaaa	gtctggtatg	tggtgccact	180
aatattgata	taaaagcttaa	ggactatgga	gtggatctta	ttgaagtctc	agacaagtga	240
tgtggggtag	agaagaagaaa	cttcgaaggc	taaaactctga	aacatcacac	atctaaagatt	300
caagagtgtg	ccgaacctaac	tcaggttgaa	acttttggct	ttcgggggga	agctctgagc	360
tcaactttgtg	cactgagcga	tgtcaccatt	tctacctgcc	acgcatcgcc	gaaggttgga	420
acttga						426

## GTBP (human) (SEQ ID NO:26)

MSRQSTLYSF	FPKSPALSDA	NKASARASRE	GGRAAAAAPGA	SPSPGGDAAW	SEAGGPRPL	60
ARSAAPPKAK	NLNGGLRRSV	APAAPTSCDF	SPGDLVWAKM	EGYPWWFCLV	YNHFFOGTFI	120
REKGSVVRVH	VQFFDDSPTR	GWVSKRLLP	YTGSKSKEAQ	KGGHFYSAPK	EILRMQRAD	180
EALNKKIKIR	LELAVCDPES	EPEEEEMEVE	GTTYVTDKSE	EDNEIESEEE	VQPKTCGSR	240
SSRQIKRRV	ISDSESDIGG	SDVEFKPDTK	EGSSDEISS	GDGSESEGL	NSPVKVARRR	300
KRMVTNGSL	KRKSRRKETP	SATKQATSIS	SETKNTLRAF	SAPQNSSEQA	HVSGGGDDSS	360
RPTVYVHETL	EWLKEEKRRD	EHRRRPDHPD	FDASTLYVPE	FDNLNSTCPK	RMWQILSKQN	420
FDLVICYKVG	KFYELVHMDA	LIGVSELGLV	FMKNWAHSG	FPEIAGFRYS	DSNVQKRGYV	480
ARVEQTEPE	MMEARCRKMA	HISKYDRVRV	REICRIITKG	QTQYSVLEGD	PSENYSKYL	540
SLKEKEEDSS	GHTRAYGVCF	VDTSLGKFFI	GQFSDDRHCS	RFRTLVAHVP	PQVLFKEGN	600
LSKCTKITLK	SSLSCSLQEG	LIPGQFQWDA	SKTLRTLLEE	EYFREKLSDG	IGVMLPQVLK	660
GMTSESDSIG	LTPGEKSELA	LSALGCGVVF	LKCLLDQEL	LSMANFEEYI	PLSDTVSTT	720
RSGAIFTKAY	QRMVLDAVTL	NNLEIFLNGT	NGSTEGTLE	RVDTCTNPPG	KRLKQWLCA	780
PLCNHVAIND	RDAIEDLMV	VPKISEVVE	LLKKLPDLER	LLSKTHNVGS	PLSKQNHPS	840
RAIMYEETTY	SKKKIIDFLS	ALEGFKVMCK	IIGIMEEVD	GFKSILKQV	ISLQTKNPEG	900
RFPDLTVELN	RWDTPADHEK	ARKTGLITPK	AGFDSYDQA	LADIRENEQS	LLEYLEKQRN	960
RIGCRITIVY	GIGRNRYQLE	IPENFTTRNL	PEEYELKSTK	KGCKRYWTKT	TEKKLANLIN	1020
AEERRDVSLK	DCMRRLFYNF	DKNYKDWQSA	VECIAVLVDL	LCLANYSRGG	DGPMCRPVIL	1080
LPEDTPVFLE	LKGSRHPCIT	KTFGGDDFIP	NDILIGCEEE	EQENGAKYCV	LVGTGPMNGGK	1140
STLMRQAGLL	AVMAQMGCVV	PAEVCRLTPI	DRVFTRLGAS	DRMSGESTF	FVELSEATSI	1200
LMHATAHSLV	LVDELGRGTA	TFDGTALANA	VVKELAEATIK	CRTLFTSTHYH	SLVDYSQNV	1260
AVRLGHMCM	VENECDDPSQ	ETITFLYKFI	KGACPKSYGF	NAARLANLPE	EVIQKGRHRA	1320
REFEKNQSL	RLRFVCLAS	ERSTVDAEAV	HKLLTLIKEL			1360

## GTBP (human cDNA) (SEQ ID NO:27)

gccgcgcggt	agatgcgggt	cttttaggag	ctccgtccga	cagaacggtt	gggccttgcc	60
------------	------------	------------	------------	------------	------------	----

ggctgtcggtt atgtcgccgac agagcaccct gtacagcttc ttccccaagt ctccggcgct 120  
 gagtgaatgcc aacaaggcct cgcccgaggc ctcaecgcgaa ggcgcgcgtg 180  
 ccccgggggc tctccttccc caggcgggga tgcggcctgg agcggaggct ggccctgggc 240  
 caggcgctgg gcgcctctgc cgtcacgcgc caagcggaag accctcaagc ggccgctgcg 300  
 gagatcggtg gcgcctctgc ccccccaccg ttgtgacttc tcaccaggag atttggtttg 360  
 ggccaagatg gagggttacc cctgggtggc ttgtctgggt tacaacacc cttttgatgg 420  
 aacattcaat ccgcgaaaag ggaatcagtt ccgtgttcat gtacagtttt ttgatgacag 480  
 cccaacaagg ggctgggtta gcaaaaggct tttaagacca tatacagctt caaatgacaa 540  
 ggaagcccaag aaggagggtc atttttacag tgcaaaagcct gaaatactga gagcaatgca 600  
 acgtgcagat gaagccttaa ataaagacaa gattaagagg cttgaattgg cagtttgtga 660  
 tgagccctca gagccagaag aggaagaaga gatggaggta ggcacaactt acgttaacaga 720  
 taagagtga gaagataatg aaattgagag tgaagaggaa gtacagccta agacacaaagg 780  
 atctagcgca agtagccgccc aaataaaaaa acgaagggtc atacagatt ctagagagtga 840  
 catttggtggc tctgatgtgg aatttaagcc agacactaag gaggaaaggaa gctgtgatga 900  
 aataaagcagt ggagtggggg atagtgaag tgaaggcctg aacagccctg tcaaaagtgtc 960  
 tcgaagcgg aagagactaa tgactggaaa tggctctctc aaaagggaaa gctctaggaa 1020  
 ggaaaacggc tcagccacca aacaagcaac tagcatttca tcagaaaacca agaattcttt 1080  
 gagagctttc tctgcccttc aaaattctga atcccaagcc caggttgttg ggttgtgtga 1140  
 tgacagtagt cgccctactg ttgggtatca atcccaagcc caggttgttg agggagataa 1200  
 gagaagtagt gagcacagag gaggccctga tcccccagat ttgtgtgcat ctacactcta 1260  
 tgtcctcag ttgttctcca attctgtac tctcgggatg aggaagtgtg ggacagttaa 1320  
 tctccagaag tttatctctg tcatctgtta caaagggtgg aaattttatg agctgtacca 1380  
 cactgatgat cttatgtgag tcaagtgaact ggggtctgta ttcatgaag gcaactgggc 1440  
 ccattctcgc ttctctgaaa ttgcatttgg ccgttatcca gattccctgg tgcataatgg 1500  
 ctataaagta ccgcagtgat aacagactga gactccagaa atgatggagg cagcgtgtag 1560  
 aaagatggca catatatcca agtatgatag aggtgtgagg agggagatct gtaggatcat 1620  
 taccaaaggt acacagactg acagtgtgct ggaagtgtat cctctgaga actacagtaa 1680  
 gatctctctt tctcccaagg aaaaagagga agattctctt ggccatctc ttgtcatatg 1740  
 tgtgtgcttt gttgataact gtgtgatact gtttttcaat ggctcagttt cagatgatcg 1800  
 ccattgttgc agattttagc ctctagtggc ctctactccc acactacac ccagtacaag 1860  
 aaaagaaat ctctcaaaag aaactaaaac aattctaaag agttcatgtg cctgttctct 1920  
 tcaggaaagt ctgatacccg gctcccagtt ttgggagcca tccaaaactt tgagaactct 1980  
 ccttgaggaa gaatatcttt cagagtctga aagtgtgagc attggggatg ttgtacccca 2040  
 agtgacttaa ttatcaactt cagagtctga ttccattggg ttgacaccag tgagaaaaag 2100  
 tgaattggcc ctctctgctc taatttttga taggtgtgtg tgtctctcac ctcaaaaaat 2160  
 tcaggagcct ctatcaatgg agaatatatt agaataat ccaagcctat cccttgattt 2220  
 cagcactaca agatctggtg aacaccttgg ctatcttcac caaagcctat caacgaatgg 2280  
 agtgacatta acaactgttg agatttttct gaatggaaca aatggttcta cgtgaagcagc 2340  
 cctactagag aggggtgtga ctgtgccata tctttttgtt aatgggttcta agcggcgtcc 2400  
 gctttgtgccc ccaactctga accattatgc tattaatgat cgtctagctg cctatagaag 2460  
 cctcatggtt gtgcctgaca aaattccga agttgtagag ctctctaaag agcttccaga 2520  
 tcttgagagg ctactcacta aaattcataa ttgtgggtct ccctgaaga cgtacagaag 2580  
 cccagcagc agggctataa tgtatgaaga aactacatac agcaagaaga agattattga 2640  
 ttctctcttc gattcgaag gattcgaagc aatgtgtaaa attatagga ctatggaaga 2700  
 agttgtgatg gggttttaagt ctaaaatcct taagcaggtc atctctctgc acagcaaaaa 2760  
 tctgaaggt cgttttctctg atttgactgt agaattgaac cgatgggata gacgtcttga 2820  
 ccattgaaaag gctcgaaga ctggaactat taectccaaa gcaggctgtt actctgatta 2880  
 tgcaccaagt ctgtctgaca taagagaaaa tgcacagagc cctcgtgaat actctagaa 2940  
 acagcgcaac agaatgtgct taggagccat agtctatttg gggatttgtt ggaaccgtta 3000  
 ccagctggaa attcctgaga attcaccac tgcgaatttg ccagaagaat aggttgtgaa 3060  
 atctaccaag aagggtctga aacagtaact gaccaaaaat attgaaaaa ggtgtgctaa 3120  
 tctcataaat gctgaagaac ggagggtatg atcattgaag gactgcagtc cagctactgt 3180  
 ctataaactt gataaaaatt acaaggactg gcagtctgct gatagtgta ttgcagttgt 3240  
 ggaatgttt ctgtgcctgg ctactatag tgcagggggt gtgtgtccta ttgtgcctc 3300  
 agtaattctg ttgccgggaag ataccocccc ctctctagag cttaaaagat taagccatcc 3360  
 ttgcaattac ttgacacttt ttggagatga tttttatct ttttatagc aatgacattc caatgactg 3420  
 tgaggaaag gagcagaaaa atggcgaagc ctatttgtgt ctgttactg cccacaaatt 3480  
 gggggccaag tctacgctta ttgacagctg gctgtaatg gatagagtgt cccagatggg 3540  
 ttgttcactg tctcagctta ttgtcagggt tttgttgaat taagtactag 3600  
 tgggtcctca ttgcaggtga tgcacagcaca ttctctgggt attttaggaa 3660  
 tgccacgata ctcatgcatg caacagcaca

aggtactgca	acatttgatg	ggacggcaat	agcaaatgca	gttggttaag	aacttgctga	3780
gactataaaa	tgtcgtacat	tattttcaac	tcactaccat	tcacttagtag	aagattatct	3840
tcacaaatgt	gctcgtgcgc	taggacata	ggcatgcagt	gtagaaagat	aatgtgaaga	3900
ccccagccat	gagactatta	cgcttcctca	taaattcatt	aaggagagct	gtccctaaaag	3960
ctatggtctt	aaagcagcaa	ggctgtctaa	tcctccagag	gaagctattc	aaaagggaca	4020
tagaaaagca	agagaaattg	agaagatgaa	tcagtcaact	cgattattct	gggaagtttg	4080
cctggctagt	gaagaagtca	ctgttagatgc	tgaagctgtc	cataaattgc	tgaactttgat	4140
taaggaatta	tagactgact	acattgggaag	ctttgagtgt	actctgcaga	aaggtggtaa	4200
attcagacaa	cattatgatc	taataaaact	tattttttaa	aaat		4244

## MSH3 (human) (SEQ ID NO:28)

MSRRKPASGG	LAASSAPAR	QAVLSRFFQS	TGSLKSTSSS	TGAADQVDPG	AAAAAAPAP	60
AFPPPLPFHV	ATEIDRRKKR	PLENDGFVKK	KVKVQVQKEG	GSDLGMSGNS	EPKCLRLTRN	120
VSKSLEKLKE	FCCDSALPQS	RVQTESLQER	FAVLPKCTDF	DDISLHAKN	AVSSSDSKRQ	180
INQKDTTLFD	LSQFGSSNTS	HENLQKTASK	SANKRSKSIY	TPLELQYIEM	KQHKDAVLC	240
VECGKYRFF	GEDAEIAARE	LNIYCHLDHN	FMASIPTHR	LFVHVRRLVA	KGYKGVGVKQ	300
TETAALKAIG	DNRSSLSFRK	LTALYTKSTL	IGEDVNPPLK	LDDAVNVNDS	MTDTSYSLL	360
CISENKENVR	KKKKGNFIFG	IVGVQPATGE	VVDFSQDSA	SRSELET RMS	SLQPEVLLLP	420
SALSQTEAL	IHRATSVSVQ	DDRIRVERMD	NIYFEYSHAF	QAVTEFYAKD	TVDIRKSGII	480
SGIVNLEKVR	ICSLAAI1KY	LKEFNLEKML	SKPENLEKPL	SKMEFTTING	TTLRNLLEIQ	540
NOTDMKTKGS	LLWLWDLTKT	SFGRRLKKW	VTQPLLLKRE	INARLDAYSE	VLHSESSVFG	600
QIENIRKLKP	DIERGLCSYI	HKKCTQEFF	LIVTKLYHLK	SEFOAIIPAV	NSHIQSDLLR	660
TVILEIPELL	SEVEHYLKIL	NEQAAKVGDK	TELFKDLSDP	FLIKRKKDEI	QGVIDEIRMH	720
LQEIIRKLKN	PSAQYVTVSG	QEFMIEIKNS	AVSCIPTDWV	KVGSTKAVSR	FHSPFIVENY	780
RHLNQLREQL	VLDCSAEWLD	FLFKFSHYH	SLCKAVHHLA	TVDCIFSLAK	VARQGDYCRP	840
TVQERKIVL	KNGRHPVIDV	LLGEQDQYVE	NMTDLSEDE	RVMIITGNPM	KGKSYIKQV	900
ALITVMAQIG	SYVPAEATII	GIVDGIFTRM	GAADNLYKGR	STEMEELTDT	AEIIRKATSQ	960
SLVILDELGR	GTSTHDGIAI	AYATLEYFIR	DVKSLLTFVT	HYFPVCELEK	NYSHQVGNVH	1020
MGFLVSEDES	KLDPGAEEQV	PDFVTFLYQI	TRGIAARSYG	LNVAKLADV	GEILKKAHAK	1080
SKELEGLINT	KKKRLKYFAK	LWTHMNAQDL	QKWTEEFNME	ETQTSLLH		1128

## MSH3 (human DNA) (SEQ ID NO:29)

gggcacgagc	cctgcacatg	ctgcgcggaa	gctgcgctgc	ggcgccctcg	ctgcctccag	60
ctccagccctc	cgagagcgaag	cggttttgag	ccgattcttc	cagtcctaccg	gaagcctgaa	120
atccacctcc	tcctccacagc	gtgcagccga	ccaggtggac	ctcgccgctgc	cagcggccgc	180
agcgccccca	cgcccccgcct	ctcccgccca	gctgcgccgc	cacgtagaact	cagaaattga	240
cagaagaaaag	aaagagccat	tggaaaaatga	tggcgctgtg	aaaaagaaaag	taaaagaaat	300
ccaaacaaaag	gaagagggaa	gtgatctggg	aatgctctgc	aactctgagc	caaaagaaat	360
tcctgagacc	aggaatgttt	caaaagtctc	ggaaaaaattg	aaagaatgct	ctgcgcgattc	420
tgccctctct	caaaagttag	tcacagacaga	atctctgcag	gagagattgc	cagttctgcc	480
aaaatgtact	gattttgatg	atatcagctc	ctctacacga	agaagtacag	ttctctctga	540
agattcgaaa	cgtcaaatata	atcaaaagga	cacaaacact	tttgatctca	ctcagtttgg	600
atcatcaaat	acaaagtcatg	aaaattttaca	gaaaaactgct	tcacaaatcag	ctaaacaaacg	660
gtccaaaaag	atctatacgc	cgctagaatt	acaatacata	gaaatgaagc	agcagcacaa	720
agatgcaggt	ttgtgtgtgc	aatgtggata	taagtataga	ttctttgggg	aaagtgcaga	780
gattgcagcc	cgagagctca	atatattatg	ccatttagat	cacaacttta	tgacagcaag	840
tatacctact	cacagactct	ttgttcatgt	agcgcgcctg	gtggcagaag	gatataaggt	900
gggagttgtg	aagcaaaact	aaactgcagc	attaaagccc	attggagaca	acagaagtct	960
actctttttc	cggaaaaatga	ctgcctttta	tacaaaaatc	acactatttg	gagaagaattg	1020
gaatccccca	atcaaaagctg	atgatgctgt	aaatgttgat	gagataatga	ctgatacttc	1080
taccagctat	ctctctgtgca	ttctctgaaa	taaggaaaaat	gttagggaga	aaaaaaaagg	1140
caacattttt	attggcattg	tgggagtgca	gcctgccaca	ggcgaggttg	tgtttgatag	1200
tttcaggagc	ttctgctctc	gttcagagct	agaaaaccgcg	atgtcacacc	tcagccagct	1260
agagctgctg	ctctcctcgg	cctgtgccga	gcaaaacagag	gcgctcatcc	acagagccac	1320
atctgttagt	gtgcagagtg	acagaattcg	agtgcgaagg	atggataaca	ttatttttga	1380
atcacagcat	gcttttccag	cagtttacaga	gttttatgca	aaagatacac	tgtacatacaa	1440
aggttctcaa	attattttctg	gcattgttaa	cttagagaag	ctctgtgctt	gctctttggc	1500
tgccatcata	aaataacctca	agaatttcaa	cttgaaaaag	atgctctcca	aacctgcagaa	1560

ttctaaacag ctatcaagta aaatggaatt tatgacaatt aatggaacaa cattaagggaa 1620  
 ttctggaatc ctacagaatc agactgatag gaaacccaaa ggaagtgtgc tgtgggtttt 1680  
 agaccacact aaaaacttcat ttggggagacg gaagttaaag aagtgggtga cccagccact 1740  
 ccttaaaata agggaaataa atgcccggct tgatgctgta tcggaaagttc tccattcaga 1800  
 actctagtgtg ttgtgtcaga tagaaaatca tctacgtaaa ttgcccgaca tagagagggg 1860  
 actctgtagc atttatcaca aaaaatgttc tacccaagag ttcttcttga ttgtcaaaac 1920  
 ttatatcac ctaaggcgac aaattcaagc aataatacct gctgttaatt cccacattca 1980  
 gtacagactg ctccggaccg ttattttaga aattcctgaa ctctcagtc agctggagca 2040  
 ttacttaaa atactcaatg aacaagctgc caaagttggg gataaaacct aattatttaa 2100  
 agacctttct gacttccctt taataaaaaa gaggaaggat gaaattcaag gtgttattga 2160  
 cgagatccga atgcatttgc aagaaatacg aaaaatacta aaaaactcct ctgcacaata 2220  
 tgtgacagta tcaggacagg agtttatgat agaaaataag aactctgctg tatcttgtat 2280  
 accaactgat tgggtaaaag ttggaagcac aaaaagctgtg agccgcttcc actctccttt 2340  
 tattgtagaa aattacacat atctgaatca gctccgggag cagctagctc ttgactgcag 2400  
 tgctgaatgc cttgattttc tagagaaatt cagtgaacat tatcactcct tgtgtaaagc 2460  
 agtgcactgc ctagcaactg ttgactgcct ttctccctg gccaaaggtg ctaagcaagg 2520  
 agattactgc agaccaactg tacaagaaga aagaaaaatt gtaataaaaa atggaaggca 2580  
 cctctgtatt gatgtgttgc tgggagaaaca ggatcaatat gtcccaataa atacagattt 2640  
 atcagaggac tcagagagag taatgataat taccggacca aactaggggt gaaagagctc 2700  
 ctacataaa caagtgcct tgattaccat catggctcag attgctcct atgttccctc 2760  
 agaagaagcg acaattggga ttgtggatgg cattttcaca aggatgggtg ctgcagacaa 2820  
 tatataataa gcaggagga ctttatgga agaactgact gacacagcag aataaatcag 2880  
 aaaaacaaca tcacagtcct tgggttatcct ggatgaacta ggaagaggga cgagcaactca 2940  
 tgatggaatt gccattgcct atgctacact tgagtatttc atcagagatg tgaattcctt 3000  
 aaccctgttt gtcccccatt atccgccagt ttgtgaacta gaaaaaaaat actcacacca 3060  
 ggtgggggaat taccacattg gattcttggt cagtggagat gaaagcaaac tggatccagg 3120  
 cgacagagaa caagtccctg attttgtcac ctctcttac caaataacta gaggaattgc 3180  
 agcaaggagt tatggattaa atgtggctaa actagcagat gttcctggag aatttttgaa 3240  
 gaaagcagct cacaagtc aaagagctgga aggattaata aatacgaataa gaaagagact 3300  
 caagtatttt gcaaaagtat ggacgatgca taatgcacaa gacctgcaga agtggacaga 3360  
 ggagttcaac atggaagaaa cacagacttc tcttcttcat taaaatgaag actacattttg 3420  
 tgacaaaaaa atggagaatt aaaaatacca actgtacaaa ataatctccc agtaaacgcc 3480  
 tatctttgtg tgacatgtga gcataaaaat atgaccatgg tatattccta ttggaacacg 3540  
 agaggttttt ctgaagacag tctttttcaa gtttctgtc tctaacttt tctacgtata 3600  
 aacactcttg aatagacttc cactttgttaa ttgaaaaatt ttatggacag taagtccagt 3660  
 aaagccttaa gtggcagaat ataattccca agcttttgga gggatgata aaaaatttat 3720  
 tgatattttt atttgttcaa gttcagataa ttggcaactg ggtgaattct gcaggaaatt 3780  
 atccattgaa ctaaaataat ttattatgc ttgccaggca tgggtgctca tgctgtaat cccagcactt 3900  
 tttttataag tagaagaagt tggccaggga atcaactgag gtcaggagt ccaagaccag ctggcccaaca 3960  
 tgggcaaaacc ccatctttac taaaaataa aagtaacatct ctactaaaaa tacgaaaaaa 4020  
 ttagctgggc atgtgtggcg acacctgtag tccagactac tccggaggct gaggcaggag 4080  
 aatctcttga acctgggagc cggaggttgc aatgagccga gatcacgtca ctgcactcca 4140  
 gctttgggca cagagcaaga ctccatctca aaaaagaaaa agaaaaagaa atagaattat 4200  
 caagctttta aaaaactagag cacagaagga ataaagtgct gaaatttaaa aggttaataa 4260  
 ttgtcatagg attaagcagt tttaagattg ttggatgaaa ttatttgtca ttcattcaag 4320  
 taataaataa tttaatgaata cttgctataa aaaaaaaaaa aaaaaaaaaa aaaa 4374

Each reference cited herein is hereby incorporated by reference in its entirety.